STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 141465

TO: Jon E Angell
Location: REM-2C18 2d2o
Art Unit: 1635
Thursday, January 06, 2005

Case Serial Number: 09/701618

From: Noble Jarrell
Location: Biotech-Chem Library
Rem 1B71
Phone: 272-2556

Noble.jarrell@uspto.gov

Search Notes

This Page Blank (uspto)

STIC-Biotech/ChemLib

From:	Angell, Jon E
Sent:	Wednesday, December 29, 2004 11:14 AM
To:	STIC-Biotech/ChemLib
Subject:	RE: Sequence Database Search Request

My apologies for the typo...

The correct serial number is: 09/701,618 CRFE

Thanks,
Eric

-----Original Message-----
From:	STIC-Biotech/ChemLib
Sent:	Wednesday, December 29, 2004 11:11 AM
To:	Angell, Jon E
Subject:	RE: Sequence Database Search Request

There is no valid CRF for this serial number, please provide us with another valid serial number.(in our in-house database)

Thank you.
Gary

-----Original Message-----
From:	Angell, Jon E
Sent:	Tuesday, December 28, 2004 4:01 PM
To:	STIC-Biotech/ChemLib
Subject:	Sequence Database Search Request

SEARCH REQUEST FORM
Scientific and Technical Information Center

Examiner# : 78697
Art Unit : 1635
Phone Number: 571-272-0756
Date: 12/28/04
Serial Number: 09/701,608 Rothbarth et al.
Mailbox & Bldg/Room Location: REMSEN 2C18
Results Format Preferred (circle): **Paper**

**********************	**********************	**********************
STAFF USE ONLY	Type of Search	Vendors and cost where applicable
Searcher: Noble	NA Sequence: #___4___	STN:_____
Searcher Phone: 2-	AA Sequence :#___2___	DIALOG:_____
Date Searcher Picked up:	Structure: #_____	QUESTEL/ORBIT:_____
Date Completed: 1 16 04	Bibliographic:_____	LEXIS/NEXIS:_____
Searcher Prep/Rev. Time: 6	Litigation:_____	SEQUENCE SYSTEM: COMPUGEN
Online Time:_____ To	Patent Family:_____	WWW/Internet:_____
	Other:_____	Other(Specify):_____

This Page Blank (uspto)

I would like to have a standard search performed using the following SEQ. ID NOs. from application : 09/701,608

SEQ ID NO. 1 (DNA ~1156 nucleotides long)
SEQ ID NO. 2 (Polypeptide ~141 amino acids long)
SEQ ID NO. 3 (DNA ~1040 nucleotides long)
SEQ ID NO. 4 (Polypeptide ~141 amino acids long)

Please perform a standard nucleic acid search for SEQ ID NO. 1 AND SEQ ID NO. 3

Please perform a standard polypeptide search for SEQ ID NO. 2 AND SEQ ID NO. 4
AND also perform a search for any nucleic acid sequences which encode SEQ ID NO.2 and SEQ ID NO.4

NOTE: SEQ ID NO. 1 is identified as encoding SEQ ID NO. 2
        SEQ ID NO. 3 is identified as encoding SEQ ID NO. 4
        and SEQ ID NO. 2 is highly similar to SEQ ID NO. 4 (differs by 14 amino acids) and they have the same function

Thanks,
Eric

J. Eric Angell
Art Unit 1635
Office: REMSEN 2D20
mailbox: REM 2C18
571-272-0756

This Page Blank (uspto)

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 5, 2005, 12:24:08 ; Search time 2302 Seconds
(without alignments)
2896.545 Million cell updates/sec

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Title: US-09-701-618A-2
Perfect score: 715
Sequence: 1 MAGEEINEDYPVEIHEYLSA.........LWEPKSKNASKVANKGKSKS 141

Scoring table: BLOSUM62

	Xgapop	10.0	,	Xgapext	0.5
	Ygapop	10.0	,	Ygapext	0.5
	Fgapop	6.0	,	Fgapext	7.0
	Delop	6.0	,	Delext	7.0

Searched: 4526729 seqs, 2364849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model
-Q=/cgn2_1/USPTO_spool_p/US09701618/runat_05012005_085838_24435/app_query.fasta_1.654
-DB=GenEmb1 -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09701618 @CGN_1_1 5783 @runat_05012005_085838 24435 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pri:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	715	100.0	1156	6	BD221634	BD221634 Method of
2	715	100.0	1156	6	AX009361	AX009361 Sequence
3	715	100.0	1168	6	BC016284	BC016284 Homo sapi
4	715	100.0	1172	6	AX329809	AX329809 Sequence

5		715	100.0	1172	6	AX779798	AX779798 Sequence
6		715	100.0	1172	6	AX779799	AX779799 Sequence
7		715	100.0	1172	6	X95592	X95592 H.sapiens m
8		715	100.0	1186	6	BC009589	BC009589 Human DNA
9		715	100.0	1195	9	BC009584	BC009584 Sequence
10		715	100.0	1199	9	BC005235	BC005235 Homo sapi
11		699	97.8	868856	9	AL354750	AL354750 Human DNA
12		699	97.8	195426	2	AC068741	AC068741 Homo sapi
13	o	671	93.8	426	6	AX452900	AX452900 Sequence
14	o	671	93.8	426	6	AX452901	AX452901 Sequence
15		671	93.8	172079	2	AL356095	AL356095 Human DNA
16	o	667	93.3	420	6	AX467454	AX467454 Sequence
17	o	667	93.3	420	6	AX467455	AX467455 Sequence
18		657	91.9	215616	2	AC109025	AC109025 Rattus no
19		657	91.9	224919	2	AC110304	AC110304 Rattus no
20		654	91.5	1040	10	AF031426	AF031426 Mus muscu
21	o	654	91.5	2000	10	BC005436	BC005436 Mus muscu
22	o	652	91.2	789	10	AY302220	AY302220 Cricetulu
23		652	91.2	1040	6	BD221635	BD221635 Method of
24		646	90.3	1050	6	AX009363	AX009363 Sequence
25		646	90.3	1050	10	MMC1DPROT	X95591 M.musculus
26		644	90.1	420	6	AX452903	AX452903 Sequence
27	o	644	90.1	420	6	AX452904	AX452904 Sequence
28	o	578.5	80.9	457	6	CO704080	CO704080 Sequence
29		519	72.6	380	6	AX885372	AX885372 Sequence
30		519	72.6	380	6	BD024982	BD024982 Sequence
31	o	501	70.1	212115	2	AC116701	AC116701 Mus muscu
32		501	70.1	238058	2	AC114654	AC114654 Mus muscu
33		438.5	61.3	154312	2	AP001767	AP001767 Homo sapi
34		438.5	61.3	178728	2	AC087613	AC087613 Homo sapi
35		425	59.4	342	11	BV179135	BV179135 sqnm10401
36		421	58.9	728	9	CR353144	CR353144 Gallus ga
37	o	380.5	53.2	4018	9	BX647771	BX647771 Homo sapi
38	o	271	37.9	155888	9	HSM807917	HSM807917 Homo sapi
39	o	252.5	35.3	209355	2	AC079112	AC079112 Homo sapi
40	o	252.5	35.3	221055	2	AC123163	AC123163 Rattus no
41	o	244	34.1	700	11	AC097131	AC097131 Rattus no
42	o	244	34.1	232869	10	BV072361	BV072361 S21P6211
43	o	244	34.1	233365	2	AL603925	AL603925 Mouse DNA
44		161	22.5	253365	2	AC044845	AC044845 Mus muscu
45		150.5	21.0	110000	8	AY050469	AY050469 Arabidops

ALIGNMENTS

RESULT 1
BD221634

LOCUS   BD221634   1156 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Method of inducing apoptosis in cell.
ACCESSION   BD221634
VERSION   BD221634.1 GI:33031404
KEYWORDS   JP 2002517192-A/1.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1156)
  AUTHORS   Rothbarth,K., Stammer,H., Werner,D. and Nails,P.
  TITLE   Method of inducing apoptosis in cell
  JOURNAL   Patent: JP 2002517192-A 1 18-JUN-2002;
            DEUTSCHES KREBSFORSCHUNGSZENTRUM STIFTUNG DES OFFENTLICHEN RECHTS,
            PETER NAILS
COMMENT   DEUTSCHES KREBSFORSCHUNGSZENTRUM STIFTUNG DES OFFENTLICHEN RECHTS,
            PETER NAILS
  OS   Homo sapiens (human)
  PN   JP 2002517192-A/1
  PD   18-JUN-2002
  PF   02-JUN-1999 JP 2000552267
  PR   02-JUN-1999 JP 2000552267
  PR   03-JUN-1998 DE 198 24 811.3
  PI   KARSTEN ROTHBARTH, HERMANN STAMMER, DIETER WERNER, PETER NAILS PC
            C12N15/09,A61K38/00,A61K48/00,A61P35/00,C12N15/00,A61K37/02 CC
  PN   Method of inducing apoptosis in cell
  FH   Key   Location/Qualifiers
  FT   CDS   (118)..(540)
  FT   mat_peptide   (118)..(540).

Column 1

```
FEATURES             Location/Qualifiers
  source    1. 1156
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
ORIGIN
```

US-09-701-618A-2 (1-141) x BD221634 (1-1156)

```
Alignment Scores:
Pred. No.:
Score:                  6.77e-69
Percent Similarity:     715.00
Best Local Similarity:  100.00%
Query Match:            100.00%
DB:                     6

Length:         1156
Matches:        141
Conservative:   0
Mismatches:     0
Indels:         0
Gaps:           0
```

```
Qy    1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  118 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 177
Qy   21 PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  178 TTTGAGAATTCAATTGGTGCTGTGGCAGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 237
Qy   41 AsnGluLeuLeuGlnLysGlyLeuAspAspProLeuGlnLeuAlaLysGluValSerAla 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  238 AATGAGTTGTTGCAGAAGTTGGATGATCCACTTGAACAACAGCAAAAGTGGATTTTGGTTCTCTGCA 297
Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  298 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357
Qy   81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  358 CATCCAGTAAAACAGGAATTGGAAAGGATCAGAGTATATATGAACAGAGTCAAGGAAATA 417
Qy  101 ThrAspLysLysLysSerAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  418 ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCGGCCTTCAAGATTTGTAAAAAAT 477
Qy  121 AlaLeuTrpGluProLysSerLysSerAsnAlaSerLysValAlaAlaAsnLysGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  478 GCCCTCTGGGAACCAAAATCGAAAAATCCATCAAAAGTTGCCAATAAAGGAAAAAGTAAA 537
Qy  141 Ser 141
        |||
Db  538 AGT 540
```

RESULT 2

```
LOCUS       AX009361            1156 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 1 from Patent WO9963071.
ACCESSION   AX009361
VERSION     AX009361.1  GI:9996662
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Rothbarth,K., Werner,D. and Stammer,H.
  TITLE     Method for triggering apoptosis in cells
  JOURNAL   Patent: WO 9963071-A1 09-DEC-1999;
            DEUTSCHES KREBSFORSCH (DE); ROTHBARTH KARSTEN (DE); WERNER DIETER
            (DE); STAMMER HERMANN (DE)
FEATURES             Location/Qualifiers
  source    1..1156
             /organism="Homo sapiens"
             /mol_type="unassigned DNA"
             /db_xref="taxon:9606"
  CDS       118..543
             /note="unnamed protein product"
```

Column 2

```
ORIGIN
             /codon_start=1
             /protein_id="CAC07450.1"
             /db_xref="GI:9996663"
             /translation="MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
             QKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELRIRVYMNRVKEITD
             KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"
```

US-09-701-618A-2 (1-141) x AX009361 (1-1156)

```
Alignment Scores:
Pred. No.:
Score:                  6.77e-69
Percent Similarity:     715.00
Best Local Similarity:  100.00%
Query Match:            100.00%
DB:                     6

Length:         1156
Matches:        141
Conservative:   0
Mismatches:     0
Indels:         0
Gaps:           0
```

```
Qy    1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  118 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 177
Qy   21 PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  178 TTTGAGAATTCAATTGGTGCTGTGGCAGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 237
Qy   41 AsnGluLeuLeuGlnLysGlyLeuAspAspProLeuGlnLeuAlaLysGluValSerAla 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  238 AATGAGTTGTTGCAGAAGTTGGATGATCCACTTGAACAACAGCAAAAGTGGATTTTGGTTCTCTGCA 297
Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  298 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357
Qy   81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  358 CATCCAGTAAAACAGGAATTGGAAAGGATCAGAGTATATATGAACAGAGTCAAGGAAATA 417
Qy  101 ThrAspLysLysLysSerAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  418 ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCGGCCTTCAAGATTTGTAAAAAAT 477
Qy  121 AlaLeuTrpGluProLysSerLysSerAsnAlaSerLysValAlaAlaAsnLysGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  478 GCCCTCTGGGAACCAAAATCGAAAAATCCATCAAAAGTTGCCAATAAAGGAAAAAGTAAA 537
Qy  141 Ser 141
        |||
Db  538 AGT 540
```

RESULT 3

```
LOCUS       BC016284            1168 bp    mRNA    linear   PRI 29-JUN-2004
DEFINITION  Homo sapiens nuclear DNA-binding protein, mRNA (cDNA clone MGC:9308
            IMAGE:3905517), complete cds.
ACCESSION   BC016284
VERSION     BC016284.2  GI:34782850
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1168)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
```

REFERENCE   2 (bases 1 to 1168)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   On Sep 16, 2003 this sequence version replaced gi:16740844.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@cgsc.bc.ca
            Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
            Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
            Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
            Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
            Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu,
            Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
            Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie
            Schein, Asim Siddiqui, Rob Holt, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 15 Row: a Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES   Location/Qualifiers
     source   1..1168
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:9308 IMAGE:3906517"
                     /tissue_type="Uterus, leiomyosarcoma"
                     /clone_lib="NIH_MGC_71"
                     /lab_host="DH10B"
                     /note=Vector: pCMV-SPORT6"
     gene   1..1168
                     /gene="C1D"
                     /note=synonyms: MGC12261, MGC14659, SUNCOR"
                     /db_xref="LocusID:10438"
                     /db_xref="MIM:606997"
     CDS   49..474
                     /gene="C1D"
                     /codon_start=1
                     /product="nuclear DNA-binding protein"
                     /protein_id="AAH16284.1"
                     /db_xref="GI:16740845"
                     /db_xref="LocusID:10438"
                     /db_xref="MIM:606997"
                     /translation="MGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMPWVVLATQGVNPKEHPVKQSLERIRVYMRVKEITD
                     KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"
ORIGIN
Alignment Scores:

Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X.; Gibbs,R.A.,
Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE

Pred. No.:   9
Score: 715.00   Matches: 141
Percent Similarity: 100.00%   Mismatches: 0
Best Local Similarity: 100.00%   Indels: 0
Query Match: 100.00%   Gaps: 0
DB:

```
Qy     1  MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    49  ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG  108

Qy    21  PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   109  TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA  168

Qy    41  AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   169  AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA  228

Qy    61  TyrThrLeuAsnSerMetProTrpValValLeuAlaThrGlnGlyValAsnProLysGlu  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   229  TACACATTAAATTCAATGTTTTGGGTTGTTTTATTGGCAACCCAAGGAGTTAATCCTAAG  288

Qy    81  HisProValLysGlnSerLeuGluArgIleArgValTyrMetArgValLysGluIle   100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   289  CATCCAGTAAAACAGAGTTTGGAAAGAATTCGAGTATATATGAGAGTCAAGGAAATA   348

Qy   101  ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   349  ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAGAAT  408

Qy   121  AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   409  GCCCTCTGGGAACCAAAATCCAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAA  468

Qy   141  Ser  141
          |||
Db   469  AGT  471
```

LOCUS   AX329809
ACCESSION   AX329809
DEFINITION Sequence 318 from Patent WO0194629.
VERSION   AX329809.1 GI:18102787
KEYWORDS
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE   Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 318 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES   Location/Qualifiers
     source   1..1172
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.:   6
Score: 715.00   Length: 1172
Percent Similarity: 100.00%   Matches: 141
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: Gaps: 0

AX329809   Sequence 318 from Patent WO0194629.   1172 bp   DNA   linear   PAT 09-JAN-2002

Pred. No.:   6.85e-69
Score: 715.00   Length: 1168
Percent Similarity: 100.00%   Matches: 141
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: Indels: 0
                                  Gaps: 0

US-09-701-618A-2 (1-141) x AX329809 (1-1172)

```
Qy      1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
Db    118 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 177

Qy     21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
Db    178 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGAGTGTCTGTTTCTCAGA 237

Qy     41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnIleAlaLysValAspLeuValLysSerAla 60
Db    238 AATCAGTTGTTGCAGAAGTTGGAATCCACTTGAACAACAAAAGTGGATTGGTTTGTTCTGCA 297

Qy     61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAlaAsnProLysGlu 80
Db    298 TACACATTAAATTCAAATGTTTTGGGTTTATTGGCAACCCAAGGAGTTAATCCTAAGGAA 357

Qy     81 HisProValLysGlnLeuGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
Db    358 CATCCAGTAAAACAGGAATTGGAAGAGATATATAGGACAGAGTGTCAAGGAAATA 417

Qy    101 ThrAspLysLysLysGlyAlaGlyValTyrAlaAlaSerArgPheValLysAsn 120
Db    418 ACAGACAAGAAAAGGCGTGGCAAGCTGACAGAGGTGCCAGCTTCAAGATTTGTAAAAAAT 477

Qy    121 AlaLeuTrpGluProLysSerLysValAlaAsnLysGlyLysSerIleLys 140
Db    478 GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAAGTATA 537

Qy    141 Ser 141
Db    538 AGT 540
```

Alignment Scores:
Pred. No.:
Score: 6.88e-69
Percent Similarity: 715.00   Length: 1172
Best Local Similarity: 100.00%   Matches: 141
Query Match: 100.00%   Conservative: 0
DB: Mismatches: 0
                                            Indels: 0
                                            Gaps: 0

US-09-701-618A-2 (1-141) x AX779798 (1-1172)

```
Qy      1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
Db    118 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 177

Qy     21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
Db    178 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGAGTGTCTGTTTCTCAGA 237
```

US-09-701-618a-2.rge

Alignment Scores:
Pred. No.:
Score: 6.88e-69
Percent Similarity: 715.00   Length: 1172
Best Local Similarity: 100.00%   Matches: 141
Query Match: 100.00%   Conservative: 0
DB: Mismatches: 0
                                            Indels: 0
                                            Gaps: 0

US-09-701-618A-2 (1-141) x AX779799 (1-1172)

```
Qy      1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
Db    118 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 177

Qy     21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
Db    178 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGAGTGTCTGTTTCTCAGA 237
```

RESULT 7

```
HSC1DPROT
LOCUS       HSC1DPROT                1172 bp    mRNA     linear   PRI 20-FEB-1998
DEFINITION  H.sapiens mRNA for C1D protein.
ACCESSION   X95592
VERSION     X95592.1  GI:1185118
KEYWORDS    C1D gene; C1D protein; DNA-binding protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nehls,P., Keck,T., Greferath,R., Spiess,E., Glaser,T.,
            Rothbarth,K., Stammer,H. and Werner,D.
  TITLE     cDNA cloning, recombinant expression and characterization of
            polypeptides with exceptional DNA affinity
  JOURNAL   Nucleic Acids Res. 26 (5), 1160-1166 (1998)
   MEDLINE  98136194
   PUBMED   9469821
REFERENCE   2  (bases 1 to 1172)
  AUTHORS   Werner,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-1996) D. Werner, Dt. Krebsforschungszentrum,
            Biochemistry of the Cell, Im Neuenheimer Feld 280-0225, D-69120
            Heidelberg, FRG
FEATURES             Location/Qualifiers
     source          1..1172
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="C1D-4"
                     /tissue_type="term placenta"
                     /clone_lib="lambda gt10 (OGi)"
     gene            1..1172
                     /gene="C1D"
     CDS             118..543
                     /gene="C1D"
                     /note="DNA-binding protein"
                     /codon_start=1
                     /product="C1D protein"
                     /protein_id="CAA64845.1"
                     /db_xref="GI:1185119"
                     /db_xref="GOA:Q13901"
                     /db_xref="TREMBL:Q13901"
                     /translation="MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVYLIATQGVNPKEHPVKQELERIRVYMNRVKEITD
                     KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"
     polyA_signal    1050..1055
```

```
ORIGIN      /gene="C1D"
Alignment Scores:
Pred. No.:              6.88e-69
Score:                 715.00
Percent Similarity:    100.00%
Best Local Similarity: 100.00%
Query Match:             9
DB:

US-09-701-618A-2 (1-141) x HSC1DPROT (1-1172)
```

```
Qy     41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 297

Qy     61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357

Qy     81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTGTATATGAACAGAGTCAAGGAAATA 417

Qy    101 ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 ACAGACAAGAAAAAGGCTGGCAAGCTGGATCGTGGAGCAGCGAGCAGATTTGTAAAGAAT 477

Qy    121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAA 537

Qy    141 Ser 141
          |||
Db    538 AGT 540
```

RESULT 8

```
BC009589
LOCUS       BC009589                1186 bp    mRNA     linear   PRI 29-JUN-2004
DEFINITION  Homo sapiens nuclear DNA-binding protein, transcript variant 2,
            mRNA (cDNA clone MGC:14659 IMAGE:4102498), complete cds.
ACCESSION   BC009589
VERSION     BC009589.1  GI:16307025
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1186)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
```

```
Qy      1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 177

Qy     21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 TTTGAGAATTCCATTGGTGTGCTGTGGATGATGCTGAAGACCATGATGTCTCGTTTCTAGA 237

Qy     41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 297

Qy     61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357

Qy     81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTGTATATGAACAGAGTCAAGGAAATA 417

Qy    101 ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 ACAGACAAGAAAAAGGCTGGCAAGCTGGATCGTGGAGCAGCGAGCAGATTTGTAAAGAAT 477

Qy    121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAA 537

Qy    141 Ser 141
          |||
Db    538 AGT 540
```

REFERENCE
AUTHORS   Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
           Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
           Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE   Generation and initial analysis of more than 15,000 full-length
           human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED   12477932
REFERENCE 2 (bases 1 to 1186)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (29-JUN-2001) National Institutes of Health, Mammalian
           Gene Collection (MGC), Cancer Genomics Office, National Cancer
           Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
           USA
REMARK
COMMENT   NIH-MGC Project URL: http://mgc.nci.nih.gov
           Contact: MGC help desk
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: CLONTECH Laboratories, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Sequencing Group at the Stanford Human Genome
           Center, Stanford University School of Medicine, Stanford, CA 94305
           Web site: http://www-shgc.stanford.edu
           Contact: (Dickson, Mark) mcd@paxil.stanford.edu
           Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
           R. M.
           Clone distribution: MGC clone distribution information can be found
           through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
           Series: IRAL Plate: 21 Row: b Column: 9
           This clone was selected for full length sequencing because it
           passed the following selection criteria: matched mRNA gi: 27894372.

FEATURES   Location/Qualifiers
     source   1..1186
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:14659 IMAGE:4102498"
                     /tissue_type="Bone marrow, chronic myelogenous leukemia"
                     /clone_lib="NIH_MGC_54"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     gene   1..1186
                     /gene="CID"
                     /note="synonyms: MGC12261, MGC14659, SUNCOR"
                     /db_xref="LocusID:10438"
                     /db_xref="MIM:606997"
     CDS   50..475
                     /gene="CID"
                     /codon_start=1
                     /product="nuclear DNA-binding protein"
                     /protein_id="AAH09589.1"
                     /db_xref="GI:16307026"
                     /db_xref="LocusID:10438"
                     /db_xref="MIM:606997"
                     /translation="MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
                     QKLDPLEQAKVDLVSAYILNSMFWVYLIATQGVNPKEHPVKQELERIRVYMRVKEITD
                     KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 6.99e-69
                        715.00   Length: 1186
Percent Similarity: 100.00%   Matches: 141
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 9   Indels: 0
                                     Gaps: 0

US-09-701-618a-2 (1-141) x BC009589 (1-1186)

Qy 1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
Db 50 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 109

Qy   21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110 TTTGAGAATTCCATTGGTGCTGTGGACGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 169

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 170 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 229

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuIleAlaThrGlnGlyValAsnProLysGlu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 230 TACACATTAAATTCAATGTTTTGGGTTTATTTGATCGCAACCCAAGGAGTTAATCCTAAGGAA 289

Qy   81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetArgValLysGluIle 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 290 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGCGAGTGAAGGAAATA 349

Qy 101 ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 350 ACAGACAAGAAAAAAGCTGGGAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAT 409

Qy 121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 410 GCCCTCTGGGAACCAAAATCCAAAAATGCATCAAAAGTTGCCAATAAAGGGAAAAGTAAA 469

Qy 141 Ser 141
        |||
Db 470 AGT 472

RESULT 9
BC009584
BC009584
BC009584
LOCUS   BC009584   1195 bp   mRNA   linear   PRI 29-JUN-2004
DEFINITION Homo sapiens nuclear DNA-binding protein, transcript variant 1,
            mRNA (cDNA clone MGC:14647 IMAGE:4095552), complete cds.
ACCESSION   BC009584
VERSION   BC009584.1 GI:16307017
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1195)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED   12477932
REFERENCE   2 (bases 1 to 1195)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (29-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK
COMMENT   NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 21 Row: n Column: 3.

FEATURES   Location/Qualifiers
1. .1195
source   /organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:14647 IMAGE:4096552"
/tissue_type="Bone marrow, chronic myelogenous leukemia"
/clone_lib="NIH_MGC_54"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"

gene   1. .1195
/gene="CID"
/gene_synonyms: MGC12261, MGC14659, SUNCOR"
/note="synonyms: MGC12261, MGC14659, SUNCOR"
/db_xref="LocusID:10438"
/db_xref="MIM:606997"

CDS   56. .481
/gene="CID"
/codon_start=1
/product="nuclear DNA-binding protein"
/protein_id="AAH09584.1"
/db_xref="GI:16307018"
/db_xref="LocusID:10438"
/db_xref="MIM:606997"
/translation="MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
QKLDPLEQAKVDLVSAYTLNSMFWVLATQGVNPKEHPVKQELERIRVYMRVKEITD
KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"

ORIGIN

US-09-701-618A-2 (1-141) x BC005984 (1-1195)

Alignment Scores:
Pred. No.:   1
Score: 715.00   Length: 1195
Pred. No.:   7.05e-69   Matches: 141
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 9   Gaps: 0

1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 115

21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
116 TTTGAGAATTCAATTGGTGCTGTGGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 175

41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
176 AATGAGTTGTTGCAGAAGTTGGACCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 235

61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
236 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCACCACCAAGGAGTTAATCCTAAGGAA 295

81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetArgValLysGluIle 100
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
296 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAGAGTCAAGGAAATA 355

101 ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
    |||||||||||||||||||||||||||||||||||||||

Db   356 ACAGACAAGAAGAAAGGCTGGGCAGCTGGAGCTGGAGCAGGAGGTGCCAGCTTCAAGATTTGTAAAAAAT 415

Qy   121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140

Db   416 GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAGTTGCCAATAAAAGGAAAAAGTAAA 475

Qy   141 Ser 141

Db   476 AGT 478

RESULT 10
LOCUS   BC005235   1199 bp   mRNA   linear   PRI 29-JUN-2004
DEFINITION   Homo sapiens nuclear DNA-binding protein, transcript variant 1,
mRNA (cDNA clone MGC:12261 IMAGE:3936648), complete cds.
ACCESSION   BC005235
VERSION   BC005235.1   GI:13528872
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1199)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE   Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED   12477932
REFERENCE   2 (bases 1 to 1199)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (27-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 16 Row: 1 Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 27894371.

FEATURES   Location/Qualifiers
source   1. .1199

```
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="MGC:12261 IMAGE:3930648"
                    /tissue_type="Lung, mucoepidermoid carcinoma"
                    /clone_lib="NIH_MGC_59"
                    /lab_host="DH10B"
                    /note="Vector: pDNR-LIB"
     gene            1..1199
                    /gene="CID"
                    /note="synonyms: MGC12261, MGC14659, SUNCOR"
                    /db_xref="LocusID:10438"
                    /db_xref="MIM:606997"
     CDS             64..489
                    /gene="CID"
                    /codon_start=1
                    /product="nuclear DNA-binding protein"
                    /protein_id="AAH05235.1"
                    /db_xref="GI:13528873"
                    /db_xref="LocusID:10438"
                    /db_xref="MIM:606997"
                    /translation="MAGEINEDYPVIIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
                    QKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITD
                    KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"
ORIGIN

US-09-701-618a-2 (1-141) x BC005235 (1-1199)

Alignment Scores:
Pred. No.:            7.08e-69                    Length:        1199
Score:               715.00                       Matches:        141
Percent Similarity:  100.00%                      Conservative:     0
Percent Identity:    100.00%                       Mismatches:       0
Best Local Similarity: 100.00%                    Indels:           0
Query Match:           9                          Gaps:             0
DB:

Qy   1 MetAlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 123

Qy  21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 183

Qy  41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 243

Qy  61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 303

Qy  81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetArgValLysGluIle 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304 CATCCAGTAAAACAGGAATTGGAAAGAATACGAGTATATATGAGAGTCAAGGAAATA 363

Qy 101 ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 364 ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAT 423

Qy 121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 424 GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAAGCAAA 483

Qy 141 Ser 141
         |||
Db 484 AGT 486
```

```
RESULT 11
LOCUS       AL354750      186856 bp    DNA     linear   PRI 03-DEC-2001
AL354750/c
DEFINITION  Human DNA sequence from clone RP11-19501 on chromosome 10 Contains
            two novel (pseudo)genes, the 5' end of a gene for a novel protein
            and 2 CpG islands, complete sequence.
ACCESSION   AL354750
VERSION     AL354750.12  GI:10715924
KEYWORDS    HTG; CpG island; transport protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186856)
  AUTHORS   Lawlor,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 6, 2000 this sequence version replaced gi:10129439.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations,
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            This sequence is the entire insert of clone RP11-19501 The true
            left end of clone RP11-433G10 is at 166973 in this sequence. The
            true right end of clone RP11-135A24 is at 6540 in this sequence.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP11-19501 is from
            the library RPCI-11.1 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..186856
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-19501"
                     /clone_lib="RPCI-11.1"
     misc_feature    3..448
                     /note="match: GSS: Em:AQ520457"
     misc_feature    7..461
                     /note="match: GSS: Em:AQ766212"
     misc_feature    7..448
                     /note="match: GSS: Em:AQ379747"
     misc_feature    115..429
                     /note="match: GSS: Em:AQ561548"
     misc_feature    2182..2217
                     /note="match: GSS: Em:AQ413399"
     repeat_region   2182..2217
                     /note="18 copies 2 mer"
     repeat_region   3383..3408
                     /note="13 copies 2 mer tg 91% conserved"
     repeat_region   5428..5511
                     /note="42 copies 2 mer ta 82% conserved"
     misc_feature    complement(5845..6532)
                     /note="match: GSS: Em:AQ151771"
     misc_feature    complement(6088..6517)
                     /note="match: GSS: Em:AQ151771"
     misc_feature    complement(6171..6534)
                     /note="match: GSS: Em:AQ147832"
     misc_feature    6545..7024
                     /note="match: GSS: Em:AQ147832"
```

```
misc_feature      6588..7012
                  /note="match: GSS: Em:AQ436650"
repeat_region     10015..10842
                  /note="match: GSS: Em:AQ324035"
repeat_region     11148..12383
                  /note="L1PA14 repeat: matches 5314..6149 of consensus"
repeat_region     12901..13700
                  /note="L1PA14 repeat: matches 4086..5314 of consensus"
misc_feature      17708..18165
                  /note="match: GSS: Em:AQ435928"
misc_feature      18643..19068
                  /note="match: STS: Em:G52882"
repeat_region     25943..26064
                  /note="match: GSS: Em:AQ202290"
repeat_region     32371..32545
                  /note="L1PB3 repeat: matches 6026..6147 of consensus"
repeat_region     36682..38127
                  /note="L1M3c repeat: matches 1140..1315 of consensus"
repeat_region     38145..42839
                  /note="L1MA8 repeat: matches 4783..6291 of consensus"
repeat_region     43906..44653
                  /note="L1PA7 repeat: matches 1433..6141 of consensus"
misc_feature      44803..46255
                  /note="CpG island"
repeat_region     49657..50116
                  /note="L1PA7 repeat: matches 9..1460 of consensus"
gene              complement(54351..55469)
                  /gene="bA195O1.1"
mRNA              complement(54351..>55469)
                  /gene="bA195O1.1"
                  /product="bA195O1.1 (novel protein similar to CID
                  DNA-binding protein)"
                  /note="match: cDNAs: Em:X95592 Em:X95591
                  match: ESTs: Em:AA831747 Em:T83118 Em:AA352554 Em:AI004680
                  Em:AA442782 Em:AA305785 Em:AI804147 Em:AI809071
                  Em:AI000846 Em:AI351827 Em:AA805413 Em:AA859937 Em:N40848
                  Em:AI553826 Em:AA458588 Em:AI831757 Em:AA42131 Em:AI935349
                  Em:N71224 Em:AW118590 Em:AW169841 Em:AA431129 Em:M87910
                  Em:AI989469 Em:AA424253 Em:AW024957 Em:AI699498
                  Em:AI243111 Em:AA424184 Em:AA610127 Em:AI123353
                  Em:AI800532 Em:AI350721 Em:AA460755 Em:R68158 Em:M87910
                  Em:AI274315 Em:AA424184 Em:AI375251 Em:AA861428
                  Em:AI301876 Em:AA476674 Em:AI816911 Em:AI027628
                  Em:AI176155 Em:AI435863 Em:AI458192 Em:AW104208
                  Em:AW340622 Em:AI379715 Em:W01341 Em:AW924440"
                  /evidence=not_experimental
CDS               complement(55051..55386)
                  /gene="bA195O1.1"
                  /note="match: GSS: Em:AQ461100"
misc_feature      54386..54689
                  /note="match: STS: Em:G23952"
polyA_signal      54999..55406
                  /note="match: proteins: Tr:Q13901"
polyA_site        54383..54388
                  /gene="bA195O1.1"
misc_feature      55412..55487
                  /note="possible pseudogene
                  match: proteins: Tr:Q13901"
                  /product="bA195O1.1 (novel protein similar to CID
                  DNA-binding protein)"
                  /codon_start=1
                  /db_xref="GI:17364063"
                  /protein_id="CAD13204.1"
                  /translation="MLKTTMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQ
                  GVNPKKHPVKGELERIRVYMNRVKEITDKKAGKLDRGAASRFVKNALWEPKPKNASK
                  VANKGKSKS"
```

```
repeat_region     61512..61611
                  /note="match: GSS: Em:AQ463761 Em:AQ505503"
repeat_region     68837..68982
                  /note="50 copies 2 mer aa 61% conserved"
repeat_region     71617..71998
                  /note="L1PA5 repeat: matches 5997..6143 of consensus"
repeat_region     72363..72506
                  /note="L1PB3 repeat: matches 5769..6150 of consensus"
misc_feature      78102..78408
                  /note="72 copies 2 mer ta 77% conserved"
misc_feature      complement(78524..78950)
                  /note="match: GSS: Em:AQ628574"
repeat_region     78659..78758
                  /note="match: GSS: Em:AQ401519"
repeat_region     79868..80242
                  /note="L1M3b repeat: matches 102..199 of consensus"
repeat_region     80250..84310
                  /note="LTR38 repeat: matches 129..503 of consensus"
repeat_region     84306..84837
                  /note="L1PA15 repeat: matches 2098..6155 of consensus"
gene              join(87084..87121,87983..88037,109294..109356,
                  111406..111627,115595..115629,118165..118302,
                  126770..127043,127968..128040,177990..>178100)
                  /gene="bA479G22.1"
mRNA              join(87084..87121,87983..88037,109294..109356,
                  111406..111627,115595..115629,118165..118302,
                  126770..127043,127968..128040,177990..>178100)
                  /gene="bA479G22.1"
                  /product="isoform 1"
                  /note="continues as bA195O1.3 (novel protein)"
                  /evidence=not_experimental
                  match: cDNAs: Em:AK023093
                  match: ESTs: Em:AA421130 Em:AA609088"
                  match: cDNAs: Em:AK023093
                  match: ESTs: Em:AA421130 in Em:AL356053
                  /product="isoform 2"
                  /note="continues as bA453G10.1 in Em:AL356053
                  /product="bA195O1.3 (novel protein)"
misc_feature      complement(87095..87660)
                  /note="match: GSS: Em:AQ474468"
repeat_region     88174..88833
                  /note="21 copies 2 mer tg 83% conserved"
misc_feature      complement(88502..88909)
                  /note="match: GSS: Em:AQ129993"
```

Alignment Scores:

Pred. No.:	2.33e-64
Score:	699.00
Percent Similarity:	98.57%
Best Local Similarity:	98.57%
Query Match:	97.76%
DB:	9

Length:	186856
Matches:	138
Conservative:	2
Mismatches:	2
Indels:	0
Gaps:	0

US-09-701-618A-2 (1-141) x AL354750 (1-186856)

```
Qy      2 AlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAlaPhe   21
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55473 GCAGGTGAAGAAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTGTCAGCGTTT 55414

Qy     22 GluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArgAsn   41
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55413 GAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCACGATGTCTGTTTCTAGAAAT 55354

Qy     42 GluLeuLeuGlnLysGluAspProLeuGluGlnAlaLysValAspLeuValSerAlaTyr   61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55353 GAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCATAC 55294
```

RESULT 12

```
Qy    62  ThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGluHis  81
              | | | | | | | | | | | | | | | | | | | | |
Db  55293  ACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACAT  55234

Qy    82  ProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIleThr  101
              | | | | | | | | | | | | | | | | | |
Db  55233  CCAGTAAAACAGGAATTCGAAAGAATCAGAGTATATAAGACAGAGTCAAGGAAATAACA  55174

Qy   102  AspLysLysLysGlyAlaGlyLeuAspArgGlyValAlaAlaSerArgPheValLysAla  121
              | | | | | | | | | | | | | | | | | |
Db  55173  GACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAATGCC  55114

Qy   122  LeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLysSer  141
              | | | | | | | | | | | | | | | | |
Db  55113  CTCTGGGAACCAAAACCGAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAGGTAAAAGT  55054
```

LOCUS   AC068741/c
DEFINITION AC068741
            195426 bp   DNA   linear   HTG 22-NOV-2000
            Homo sapiens clone RP11-777E12, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC068741
VERSION   AC068741.3 GI:11276256
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 195426)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens, clone RP11-777E12
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 195426)

AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A. Burkett,G.,
            Campopiano,A., Castle,A. Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N. Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R. Landers,T. Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G. Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H. O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M. Oliver,J. Peterson,K., Pierre,N.,
            Pisani,C. Pollara,V., Raymond,C., Riley,R., Rogov,P. Rothman,D.,
            Roy,A., Santos,R., Schauer,S. Severy,P., Spencer,B.,
            Stange-Thomann,N. Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S. Theodore,J., Tirrell,A., Travers,M. Trigilio,J.,
            Vassiliev,H. Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (08-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Nov 22, 2000 this sequence version replaced gi:9369524.
            All repeats were identified using RepeatMasker.
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------------------- Project Information
            Center project name: L8655
            Center clone name: 777_E.12
            ---------------------- Summary Statistics
            Sequencing vector: M13; M77815; 30% of reads
            Sequencing vector: Plasmid; n/a; 70% of reads

FEATURES   Location/Qualifiers
     source   1..195426
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-777E12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1..731
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature   832..3248
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:left"
     misc_feature   3349..8620
                     /note="assembly_fragment"
     misc_feature   8721..17691
                     /note="assembly_fragment"
     misc_feature   17792..25774
                     /note="assembly_fragment"
     misc_feature   25875..66476
                     /note="assembly_fragment"
     misc_feature   66577..128195
                     /note="assembly_fragment
     misc_feature   128296..195426
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN
```

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191262 bases at least Q40
Consensus quality: 193601 bases at least Q30
Consensus quality: 194390 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 194726; sum-of-contigs
Quality coverage: 6.9 in Q20 bases; agarose-fp
Quality coverage: 6.5 in Q20 b.

NOTE: This is a 'working draft' sequence. It currently
consists of 8 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```
1      731:   contig of 731 bp in length.
732    831:   gap of 100 bp
832    3248:  contig of 2417 bp in length
3249   3348:  gap of 100 bp
3349   8620:  contig of 5272 bp in length
8621   8720:  gap of 100 bp
8721   17691: contig of 8971 bp in length
17692  17791: gap of 100 bp
17792  25774: contig of 7983 bp in length
25775  25874: gap of 100 bp
25875  66476: contig of 40602 bp in length
66477  66576: gap of 100 bp
66577  128195: contig of 61619 bp in length
128196 128295: gap of 100 bp
128296 195426: contig of 67131 bp in length.
```

ORIGIN

Alignment Scores:
Pred. No.:
Score: 2.47e-64
Percent Similarity: 699.00
Best Local Similarity: 98.57%
Query Match: 98.57%
DB: 97.76%

```
Length:          195426
Matches:         138
Mismatches:      2
Conservative:    0
Indels:          0
Gaps:            0
```

```
Qy      2  AlaGlyGluGluIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAlaPhe  21
              | | | | | | | | | | | | | | | | | | | | |
```

RESULT 13

```
LOCUS       AX452900                426 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 1 from Patent WO0242322.
ACCESSION   AX452900
VERSION     AX452900.1  GI:21712537
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Jackson,D., Casari,G. and Suckow,J.
  TITLE     Mammalian nuclear receptor cofactors cf7 and cf8 and methods of use
  JOURNAL   Patent: WO 0242322-A 1 30-MAY-2002;
            LION Bioscience AG (DE)
FEATURES             Location/Qualifiers
     source          1..426
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

US-09-701-618A-2 (1-141) x AX452900 (1-426)

Alignment Scores:
Pred. No.:
Score:               671.00
Percent Similarity:   95.74%
Best Local Similarity: 93.62%
Query Match:          93.85%
DB:                   6

1.3e-64              Length:    426
                     Matches:   132
                     Conservative: 3
                     Mismatches:   6
                     Indels:       0
                     Gaps:         0
                     1. .426
```

```
Qy    1 MetAlaGlyGluGlnIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAGATTCACGAGTATTTGTCAGCA 60

Qy   21 PheGluAsnSerIleGlyValAlaValAspGluLeuLysThrMetSerValSerArg 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTTGAGAATTCCATTGGTGTGGCTGTGGATGAGCTGAAGACAATGAGTGTCTGTTTCAGA 120

Qy   41 AsnGluLeuGlnGlyLysValAlaAspLeuGluIle 60
         ||||||||||||||||||||||||||||||||||
Db  121 AATGAGTTGTTGCAGGGAAAGGTGGCAGAGTTGGA 180
```

RESULT 14

```
LOCUS       AX452901/c              426 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 2 from Patent WO0242322.
ACCESSION   AX452901
VERSION     AX452901.1  GI:21712538
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Jackson,D., Casari,G. and Suckow,J.
  TITLE     Mammalian nuclear receptor cofactors cf7 and cf8 and methods of use
  JOURNAL   Patent: WO 0242322-A 2 30-MAY-2002;
            LION Bioscience AG (DE)
FEATURES             Location/Qualifiers
     source          1..426
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

US-09-701-618A-2 (1-141) x AX452901 (1-426)

Alignment Scores:
Pred. No.:
Score:               671.00
Percent Similarity:   95.74%
Best Local Similarity: 93.62%
Query Match:          93.85%
DB:                   6

1.3e-64              Length:    426
                     Matches:   132
                     Conservative: 3
                     Mismatches:   6
                     Indels:       0
                     Gaps:         0
                     1. .426
```

```
Db          126 ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCCAGCTTCAAGATTTGTAAGAAAT 67
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy          121 AlaLeuThrGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyGlySerLys 140

Db
Qy           66 GCCCTCTGGGAACCAAAAACCGAAAAAATACATCCAAAGTTGCCCATAAAGGAAAAAGTAAA
                ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Qy          141 Ser 141
                |||
Db            6 AGT 4
```

RESULT 15

```
LOCUS       AL356095/c                 172079 bp    DNA     linear   PRI 24-JAN-2001
DEFINITION  Human DNA sequence from clone RP11-369J21 on chromosome 10,
            complete sequence.
ACCESSION   AL356095
VERSION     AL356095.11  GI:12331029
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172079)
  AUTHORS   Brown,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 22, 2001 this sequence version replaced gi:11863400.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-369J21 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong.  For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-369J21 The true
            right end of clone RP11-479O17 is at 65134 in this sequence.
FEATURES             Location/Qualifiers
     source          1..172079
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-369J21"
                     /clone_lib="RPCI-11.2"
     misc_feature    7..431
                     /note="match: GSS: Em:AQ191918"
     misc_feature    25..413
                     /note="match: GSS: Em:AQ815029"
     misc_feature    61..595
                     /note="match: GSS: Em:AQ531010"
     repeat_region   647..944
                     /note="L1PA12 repeat: matches -1422. .-1118 of consensus"
     repeat_region   1339..1734
     misc_feature
                     /note="L2 repeat: matches 1652. .2042 of consensus"
```

```
     repeat_region   2037..2348
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   2389..2697
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   2939..3074
                     /note="AluSx repeat: matches 2. .310 of consensus"
     repeat_region   4768..5076
                     /note="FLAM_C repeat: matches 12. .141 of consensus"
     repeat_region   3197..3411
                     /note="MIR58A repeat: matches 1. .224 of consensus"
     repeat_region   3518..3821
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   5994..6157
                     /note="L1MCb repeat: matches 102. .831 of consensus"
     repeat_region   5475..5997
                     /note="AluSx repeat: matches 12. .141 of consensus"
     repeat_region   6158..6294
                     /note="L1M4 repeat: matches 4140. .4298 of consensus"
     repeat_region   4024..4580
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   4626..4766
                     /note="L2 repeat: matches 2210. .2750 of consensus"
     repeat_region   6299..6750
                     /note="FLAM_C repeat: matches 1. .133 of consensus"
     repeat_region   6751..7063
                     /note="L1M4 repeat: matches 4298. .4741 of consensus"
     repeat_region   7064..7628
                     /note="L2 repeat: matches 2240. .2383 of consensus"
     repeat_region   7629..7915
                     /note="L1M4 repeat: matches 4741. .5310 of consensus"
     repeat_region   7916..8047
                     /note="AluSx repeat: matches 1. .284 of consensus"
     repeat_region   8054..8180
                     /note="L1M4 repeat: matches 5310. .5437 of consensus"
     repeat_region   8182..8411
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   8412..8717
                     /note="FLAM_C repeat: matches 2. .140 of consensus"
     repeat_region   8718..8862
                     /note="L1PA7 repeat: matches 5432. .5661 of consensus"
     repeat_region   8863..9012
                     /note="AluSq repeat: matches 1. .306 of consensus"
     repeat_region   9140..9288
                     /note="L1PA7 repeat: matches 5661. .5805 of consensus"
     repeat_region   9292..9516
                     /note="L1PA7 repeat: matches 5992. .6141 of consensus"
     repeat_region   10088..10389
                     /note="AluJo repeat: matches 1. .151 of consensus"
     repeat_region   10390..10697
                     /note="L1MB6 repeat: matches 5971. .6182 of consensus"
     repeat_region   11104..11317
                     /note="AluSq repeat: matches 1. .302 of consensus"
     repeat_region   11749..12047
                     /note="AluY repeat: matches 1. .308 of consensus"
     repeat_region   12418..12472
                     /note="MER46A repeat: matches 22. .235 of consensus"
     repeat_region   12473..12772
                     /note="AluSq repeat: matches 1. .299 of consensus"
     repeat_region   12773..13104
                     /note="L1MB2 repeat: matches 3968. .4023 of consensus"
     repeat_region   13105..13408
                     /note="AluSg repeat: matches 1. .299 of consensus"
     repeat_region   13409..13722
                     /note="L1MB2 repeat: matches 4023. .4362 of consensus"
     repeat_region   13410..14039
                     /note="AluY repeat: matches 1. .310 of consensus"
     repeat_region   14040..14344
                     /note="L1MB2 repeat: matches 4362. .4638 of consensus"
     repeat_region   14345..14683
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   14745..15005
                     /note="L1MB2 repeat: matches 4638. .4985 of consensus"
     misc_feature    complement(14996..15536)
                     /note="AluSp repeat: matches 1. .291 of consensus"
```

DB: 9
US-09-701-618A-2 (1-141) x AL356095 (1-172079)   Gaps: 0

```
Qy      1 MetAlaGlyGluGluLeuAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla  20
            |||||||||||||||||||||||||||||||||||||||||||:|:||||||||||||||
Db  36337 ATGGCAGCTGAAGAAAATTAATGAGGAGACTATCCAGTAGGAAAATTCACGATTATTTGTCAGCA 36278

Qy     21 PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetSerValSerArg  40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36277 TTTGCGAATTCCATTCAATGCGTGTGGATGAGATGAGAGACTGAAGAACATGATGAGTCTGTTTCTAGA 36218

Qy     41 AsnGluLeuLeuGlnIleLysSerLeuAspProLeuGluGlnAlaLysValValSerAla  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36217 AATGAGTTGTTGTCAGAAGTTGGACCCCACTCAACAAGCAAAAGTGGAATTTGGTTTCTGCA 36158

Qy     61 TyrThrLeuAsnSerMetPheThrValTyrLeuAlaThrGlnGlyValAsnProLysGlu  80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36157 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAGGAGGTGAATCCTAAGGAA 36098

Qy     81 HisProValLysGlnLeuLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36097 CATCCAGTAAAGCAGCAGAATTGGAAAGAATCAGAGTATATATGGCAGTCAAGGAGAATA 36038

Qy    101 ThrAspLysLysLysAlaGlyLysLysLeuAspArgGlyValAlaAlaSerArgPheValLysAsn 120
            |||||||||||||||||||||||||||||:||:||||||||||:::|||
Db  36037 ACAGACAAGAAAAAAGGCTGGGCAAGCTGGACAGAGGGTCCAGCTTCAAGAATTTGTAAGAAT 35978

Qy    121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAlaAsnLysGlyLysSerLys 140
            |||||||||||||||||||||||||||:::|||||||||||||
Db  35977 GCCCCTCGGGGAACCAAAAACCGAAAAATACATCCAAAGTTGCCCATAAAGGAAAAAAGTAAA 35918

Qy    141 Ser 141
            |||
Db  35917 AGT 35915
```

Search completed: January 5, 2005, 14:48:12
Job time : 2401 secs

Alignment Scores:
Pred. No.:
Score: 671.00
Percent Similarity: 95.74%
Best Local Similarity: 93.62%
Query Match: 93.85%

2.48e-61
Length: 172079
Matches: 132
Conservative: 3
Mismatches: 6
Indels: 0

Feature	Location	Note
misc_feature	15137. .15273	/note="match: GSS: Em:AQ536820"
repeat_region	15310. .15539	/note="AluJo repeat: matches 1. .134 of consensus"
repeat_region	complement(15014. .15524)	/note="match: GSS: Em:AQ192026"
repeat_region	15540. .15828	/note="AluJo repeat: matches 1. .134 of consensus"
repeat_region	15829. .16547	/note="L1MB2 repeat: matches 3. .291 of consensus"
repeat_region	16558. .16666	/note="AluJb repeat: matches 5042. .5257 of consensus"
repeat_region	16675. .16979	/note="L1MB2 repeat: matches 5257. .6013 of consensus"
repeat_region	16994. .17094	/note="MLT1B repeat: matches 2. .121 of consensus"
misc_feature	complement(17009. .17406)	/note="AluSx repeat: matches 1. .306 of consensus"
repeat_region	17971. .18031	/note="match: GSS: Em:AQ076012"
repeat_region	18034. .18072	/note="L1MB2 repeat: matches 6069. .6168 of consensus"
repeat_region	18073. .18375	/note="Alu repeat: matches 2. .62 of consensus"
repeat_region	18376. .18422	/note="MER4D repeat: matches 564. .601 of consensus"
repeat_region	18490. .18634	/note="AluJb repeat: matches 2. .301 of consensus"
repeat_region	18635. .18895	/note="MER4D repeat: matches 601. .651 of consensus"
repeat_region	18947. .19130	/note="LTR29 repeat: matches 53. .198 of consensus"
repeat_region	19128. .19305	/note="MER41B repeat: matches 1. .271 of consensus"
repeat_region	19306. .19377	/note="AluSg/x repeat: matches 109. .292 of consensus"
repeat_region	19723. .19808	/note="AluSq repeat: matches 140. .313 of consensus"
repeat_region	19826. .19975	/note="36 copies 2 mer aa 75% conserved"
repeat_region	19976. .20597	/note="LTR8 repeat: matches 1. .73 of consensus"
misc_feature	complement(20789. .21658)	/note="AluJb repeat: matches 137. .286 of consensus"
repeat_region	20900. .21045	/note="LTR8 repeat: matches 60. .689 of consensus"
repeat_region	21046. .21335	/note="match: GSS: Em:AQ891390"
repeat_region	21336. .21501	/note="AluJo repeat: matches 1. .127 of consensus"
repeat_region	22184. .22478	/note="AluSx repeat: matches 6. .297 of consensus"
repeat_region	22687. .22768	/note="AluJo repeat: matches 127. .299 of consensus"
repeat_region	23426. .23535	/note="AluSx repeat: matches 1. .295 of consensus"
repeat_region	23536. .23826	/note="AluS repeat: matches 2. .84 of consensus"
repeat_region	23827. .23869	/note="HERVL repeat: matches 2. .98 of consensus"
repeat_region	23932. .24283	/note="AluSg repeat: matches 1. .294 of consensus"
repeat_region	24284. .24540	/note="HERVL repeat: matches 98. .143 of consensus"
repeat_region	24541. .24787	/note="HERVL repeat: matches 4267. .4622 of consensus"
repeat_region		/note="AluSx repeat: matches 1. .294 of consensus"

This Page Blank (uspto)

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: January 4, 2005, 03:16:16

Title: US-09-701-618A-1
Perfect score: 1156
Sequence: 1 ctttccggggagactggagtc..........aggctgaaatgttgtaaaa 1156

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 2364849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                1: gb_ba:*
                2: gb_htg:*
                3: gb_in:*
                4: gb_om:*
                5: gb_ov:*
                6: gb_pat:*
                7: gb_ph:*
                8: gb_pl:*
                9: gb_pr:*
                10: gb_ro:*
                11: gb_sts:*
                12: gb_sy:*
                13: gb_un:*
                14: gb_vi:*

Search time 5271.49 Seconds (without alignments)
10370.300 Million cell updates/sec

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	1156	100.0	1156	6	BD221634	Method of
2	1156	100.0	1156	6	AX009361	Sequence
3	1145	99.0	1172	6	AX329809	Sequence
4	1145	99.0	1172	6	AX779798	Sequence
5	1145	99.0	1172	6	AX779799	Sequence
6	1145	99.0	1186	6	HSC1DPROT	
7	927.2	80.2	1186	6	X95592	H.sapiens m
8	926	80.1	1168	6	BC009589	Homo sapi
9	922.4	79.8	1195	9	BC016284	Homo sapi
10	919.8	79.6	1199	9	BC009584	Homo sapi
11	900.4	77.9	186856	2	AL354750	Human DNA
12	900.4	77.9	195426	9	AC068741	Homo sapi
13	655.2	56.7	155888	9	AC079112	Homo sapi
14	519.2	44.9	172079	9	AL356095	Human DNA
15	464	40.1	2000	10	BC005436	Mus muscu
16	458	39.6	1040	10	AF031426	Mus muscu
17	433.8	37.5	1040	6	BD221635	Method of
18	433.8	37.5	1040	6	AX009363	Sequence
19	433.8	37.5	1050	10	MMC1DPROT	
c 20	431	37.3	443	6	AX777852	Sequence
c 21	414.2	35.8	215616	2	AC109025	Sequence
c 22	414.2	35.8	224919	2	AC110304	Rattus no
c 23	400.4	34.6	426	6	AX452900	Sequence
c 24	400.4	34.6	426	6	AX452901	Sequence
c 25	394.4	34.1	420	6	AX467454	Sequence
c 26	394.4	34.1	420	6	AX467455	Sequence
c 27	389.6	33.7	420	6	AX452903	Sequence
c 28	389.6	33.7	420	6	AX452904	Sequence
c 29	389.2	33.7	420	10	AY302220	Sequence
c 30	376.8	32.6	457	6	CQ704080	Sequence
c 31	312	27.0	380	6	AX885372	Sequence
c 32	312	27.0	380	6	BD024982	Sequence
c 33	296.2	25.6	154312	9	AP001767	Homo sapi
c 34	296.2	25.6	178728	9	AC087613	Homo sapi
c 35	265.6	23.0	212115	2	AC116701	Mus muscu
c 36	265.6	23.0	238058	2	AC114654	Mus muscu
c 37	265.6	23.0	395	11	G23952	human STS w
c 38	259.4	22.4	342	11	BV171135	sqnm10401
c 39	239.4	20.7	232869	9	AC044845	Mus muscu
c 40	239.4	20.7	235365	9	AL603925	Homo sapi
c 41	239	20.7	4018	9	BX647771	Homo sapi
c 42	200.4	17.3	728	3	CR353144	Gallus ga
c 43	194.8	16.9	209355	2	AC123163	Rattus no
c 44	194.8	16.9	221055	2	AC097131	Rattus no
c 45	154.4	13.4	700	11	BV072361	S212P6211

ALIGNMENTS

RESULT 1
BD221634

LOCUS   BD221634   1156 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Method of inducing apoptosis in cell.
ACCESSION   BD221634
VERSION   BD221634.1 GI:33031404
KEYWORDS   JP 2002517192-A/1.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1156)
  AUTHORS   Rothbarth,K., Stammer,H., Werner,D. and Nails,P.
  TITLE   Method of inducing apoptosis in cell
  JOURNAL   Patent: JP 2002517192-A 1 18-JUN-2002;
            DEUTSCHES KREBSFORSCHUNGSZENTRUM STIFTUNG DES OFFENTLICHEN RECHTS,
            PETER NAILS
COMMENT   OS Homo sapiens (human)
            PN JP 2002517192-A/1
            PD 18-JUN-2002
            PF 02-JUN-1999 JP 2000552267
            PR 03-JUN-1998 DE 198 24 811.3
            PI KARSTEN ROTHBARTH, HERMANN STAMMER, DIETER WERNER, PETER NAILS PC
            C12N15/09,A61K38/00,A61K48/00,A61P35/00,C12N15/00,A61K37/02 CC
            Method of inducing apoptosis in cell
            FH Key   Location/Qualifiers
            FT CDS   (118)..(540)
            FT mat_peptide   (118)..(540).
FEATURES   Location/Qualifiers
     source   1..1156
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match   100.0%; Score 1156; DB 6; Length 1156;
Best Local Similarity   100.0%;
Matches 1156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTTCCGGGGAGACTGGAGTCGAAGGCCGTGAGTCGAGTATTTTCTPAAGCCAGTGTTTAGAGAGTA 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 CTTTCCGGGGAGACTGGAGTCGAAGGCCGTGAGTCGAGTATTTTCTPAAGCCAGTGTTTAGAGAGTA 60

Left column (alignment):

```
Qy    61 TGTGAGGCAAGAGTACCTATAGAACCCGGAGGAGGGTGAGGAGCAGAGCTGGCCATAATG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TGTGAGGCAAGAGTACCTATAGAACCCGGAGGAGGGTGAGGAGCAGAGCTGGCCATAATG 120

Qy   121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAGAATTCACGAGTATTTGTCAGCGTTT 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAGAATTCACGAGTATTTGTCAGCGTTT 180

Qy   181 GAGAATTCCATTGGTGTCGTGGATGAGATGTCTGAAGACCATGATGTCTGTTTTGTCAGAAAT 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GAGAATTCCATTGGTGTCGTGGATGAGATGTCTGAAGACCATGATGTCTGTTTTGTCAGAAAT 240

Qy   241 GAGTTGTTGCAGAAGAGTTGGATCCACTTGAACAAGCAAAGTGGGATTTGGGTTTCTGCATAC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAGTTGTTGCAGAAGAGTTGGATCCACTTGAACAAGCAAAGTGGGATTTGGGTTTCTGCATAC 300

Qy   301 ACATTAAATTCAAATGTTTTGGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACAT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACATTAAATTCAAATGTTTTGGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACAT 360

Qy   361 CCAGTAAAAACAGGAATTGGAAAGAATCAGAGTATATATATGAACAGAGTCAAGGAAATAACA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCAGTAAAAACAGGAATTGGAAAGAATCAGAGTATATATATGAACAGAGTCAAGGAAATAACA 420

Qy   421 GACAAGAAAAAAGGCTGGGACAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAATGCC 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACAAGAAAAAAGGCTGGGACAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAATGCC 480

Qy   481 CTCTGGGACCAAAATCGAAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAAGTAAAAGT 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CTCTGGGACCAAAATCGAAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAAGTAAAAGT 540

Qy   541 TAACTTTTTGGTTTTGATGTACACATATTCAAAAAGTACACATTAATATGTAATCACAGTAA 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TAACTTTTTGGTTTTGATGTACACATATTCAAAAAGTACACATTAATATGTAATCACAGTAA 600

Qy   601 TATGTAAAGCTAAAATACTTCCTCTCCAAAGATCATTATCTTTATTGATTAGCACTGAGGA 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TATGTAAAGCTAAAATACTTCCTCTCCAAAGATCATTATCTTTATTGATTAGCACTGAGGA 660

Qy   661 TTTTAACATTGTGATATATATATATTACCATCTCTTGATGAGGAGACTCTTATT 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TTTTAACATTGTGATATATATATATTACCATCTCTTGATGAGGAGACTCTTATT 720

Qy   721 TCTTTTATATAGGTCAGTCTGCAAGTACACCATTTTATATAATTTACAGCGTGAAATTTAAGTGAA 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TCTTTTATATAGGTCAGTCTGCAAGTACACCATTTTATATAATTTACAGCGTGAAATTTAAGTGAA 780

Qy   781 ATGTTCTTTGTAAACATTTGTACTATTTTAAATGAATAATGAACCTTATGTAAGGTATGGCTAT 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ATGTTCTTTGTAAACATTTGTACTATTTTAAATGAATAATGAACCTTATGTAAGGTATGGCTAT 840

Qy   841 CTGTAGGCTGAAATTATAGGTACATCTGTTTTCACTATATATATTAAGAAAGCCGTGAAT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CTGTAGGCTGAAATTATAGGTACATCTGTTTTCACTATATATATTAAGAAAGCCGTGAAT 900

Qy   901 GACTTAAAATGTTCATTTTTTCTGTATAGATAACTTTATCATGTTTTCATCATGTTTAGGAA 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GACTTAAAATGTTCATTTTTTCTGTATAGATAACTTTATCATGTTTTCATCATGTTTAGGAA 960

Qy   961 TTACTGCTTTGTTGATATTCAAAGTGTGAAACTAAAAGTTATGGTTGTACTTTAATTCT 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TTACTGCTTTGTTGATATTCAAAGTGTGAAACTAAAAGTTATGGTTGTACTTTAATTCT 1020

Qy  1021 TGGCATGTTGCCTCTATGTCCCATTTAAAATAAAATACATTCTCATTAACTTTAGGATGGG 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 TGGCATGTTGCCTCTATGTCCCATTTAAAATAAAATACATTCTCATTAACTTTAGGATGGG 1080

Qy  1081 AAATAAGGTTGTATGTTGATGGATGAATGAACTGTACTCTCAATAAAGGC 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 AAATAAGGTTGTATGTTGATGGATGAATGAACTGTACTCTCAATAAAGGC 1140
```

Right column:

```
Qy  1141 TGAAAATGTTGTAAAA 1156
         ||||||||||||||||
Db  1141 TGAAAATGTTGTAAAA 1156
```

RESULT 2

```
Query Match          100.0%;   Score 1156; DB 6;   Length 1156;
Best Local Similarity 100.0%;   Pred. No. 6.3e-197;
Matches 1156;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 CTTTCCGGGAGACTGGAGTCGAAGGCCGTGAGTATTTCTTAAGCCAGTGTTGTTAGAGAGTA 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTTTCCGGGAGACTGGAGTCGAAGGCCGTGAGTATTTCTTAAGCCAGTGTTGTTAGAGAGTA 60

Qy    61 TGTGAGGCAAGAGTACCTATAGAACCCGGAGGAGGGTGAGGAGCAGAGCTGGCCATAATG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TGTGAGGCAAGAGTACCTATAGAACCCGGAGGAGGGTGAGGAGCAGAGCTGGCCATAATG 120

Qy   121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAGAATTCACGAGTATTTGTCAGCGTTT 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAGAATTCACGAGTATTTGTCAGCGTTT 180

Qy   181 GAGAATTCCATTGGTGTCGTGGATGAGATGTCTGAAGACCATGATGTCTGTTTTGTCAGAAAT 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GAGAATTCCATTGGTGTCGTGGATGAGATGTCTGAAGACCATGATGTCTGTTTTGTCAGAAAT 240

Qy   241 GAGTTGTTGCAGAAGAGTTGGATCCACTTGAACAAGCAAAGTGGGATTTGGGTTTCTGCATAC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAGTTGTTGCAGAAGAGTTGGATCCACTTGAACAAGCAAAGTGGGATTTGGGTTTCTGCATAC 300

Qy   301 ACATTAAATTCAAATGTTTTGGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACAT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACATTAAATTCAAATGTTTTGGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACAT 360

Qy   361 CCAGTAAAAACAGGAATTGGAAAGAATCAGAGTATATATATGAACAGAGTCAAGGAAATAACA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCAGTAAAAACAGGAATTGGAAAGAATCAGAGTATATATATGAACAGAGTCAAGGAAATAACA 420

Qy   421 GACAAGAAAAAAGGCTGGGACAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAATGCC 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACAAGAAAAAAGGCTGGGACAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAATGCC 480
```

```
RESULT 3
AX329809
LOCUS       AX329809      1172 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 318 from Patent WO0194629.
ACCESSION   AX329809
VERSION     AX329809.1  GI:18102787
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 318 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1..1172
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
Query Match           99.0%;  Score 1145;  DB 6;  Length 1172;
Best Local Similarity 99.9%;  Pred. No. 5.8e-195;
Matches 1156;  Conservative 0;  Mismatches 0;  Indels 1;  Gaps 1;
```

```
Qy      1 CTTTCCGGGAGGAGGAGTCGGAGTCGGAAGGCGTGAGTGAGTATTTTCTAAGGCCAGTTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTTTCCGGGAGGAGGAGTCGGAGTCGGAAGGCGTGAGTGAGTATTTTCTAAGGCCAGTTA 60

Qy     61 TGTGAGGCAAGAGTACCTATAGAACCCGGGAGGAGGAGGCCAGAGCCAGTGGCCAATAATG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGTGAGGCAAGAGTACCTATAGAACCCGGGAGGAGGAGGCCAGAGCCAGTGGCCAATAATG 120
```

RESULT 4

```
AX779798
LOCUS       AX779798              1172 bp   DNA   linear   PAT 14-JUL-2003
DEFINITION  Sequence 1955 from Patent WO03039443.
ACCESSION   AX779798
VERSION     AX779798.1  GI:32696792
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.,
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 1955 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE);
            Ludwig-Maximilian-Universitaet Muenchen (DE); Haferlach, Torsten,
            PD Dr. Dr. (DE); Schoch, Claudia (DE); Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1..1172
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
Query Match             99.0%;  Score 1145;  DB 6;  Length 1172;
  Best Local Similarity   99.8%;
  Matches 1156;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy         1 CTTTCCGGGAGGACTGAGTCGAGGCCGTGAGTGTATTTCTAAGCCAGTGTTAGAGAGTA 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CTTTCCGGGAGGACTGAGTCGAGGCCGTGAGTGTATTTCTAAGCCAGTGTTAGAGAGTA 60

Qy        61 TGTGAGGCAAGAGTACCTATACGAACCCGGAGGAGCAGAGCTGGGCCATAATG 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 TGTGAGGCAAGAGTACCTATACGAACCCGGAGGAGCAGAGCTGGGCCATAATG 120

Qy       121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCACGGTTT 180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCACGGTTT 180

Qy       181 GAGAATTCAATGGTGTCGTGGATAGTGTGAAGACCATGATCGTCGTTTAGAGAAAT 240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GAGAATTCAATGGTGTCGTGGATAGTGTGAAGACCATGATCGTCGTTTAGAGAAAT 240

Qy       241 GAGTTGTTGCAGAAGTTGGAATCCACTTGAACAACGAAAAGTTGGAATTTGGTTTCTGCAAC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GAGTTGTTGCAGAAGTTGGAATCCACTTGAACAACGAAAAGTTGGAATTTGGTTTCTGCAAC 300

Qy       301 ACATTAAATTCAATGTTTTGGGGTTATTTGGGCCACCAAGGACTTAATCCTAAGGACAT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ACATTAAATTCAATGTTTTGGGGTTATTTGGGCCACCAAGGACTTAATCCTAAGGACAT 360

Qy       361 CCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATAGTGCAAGGAAAATAACA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATAGTGCAAGGAAAATAACA 420

Qy       421 GACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGGATTTGTAAAAATGCC 480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGGATTTGTAAAAATGCC 480
```

```
Db      1020 TTGGCATGTTGCCTCTATGTCCCATTTAAAATAAAATACATTTCTCATTAACTTTAGATGG 1079
Qy      1021 TTGGCATGTTGTGCCTCTATGTCCCATTTAAAATAAAATACATTTCTCATTAACTTTAGATGG 1080

Db      1080 GAAATAAGGTTGTATGTTGATGGATGATGAATTTGGCATGATGAATTTGGCAATAAAAGT 1139
Qy      1081 GAAATAAGGTTGTATGTTGATGGATGATGAATTTGGCATGATGAATTTGGCAATAAAAGG 1140

Db      1140 CTGAAAAAGTTGTAAAA 1156
Qy      1141 CTGAAAAAGTTGTAAAA 1157
```

RESULT 5

```
AX779799
LOCUS       AX779799              1172 bp   DNA   linear   PAT 14-JUL-2003
DEFINITION  Sequence 1956 from Patent WO03039443.
ACCESSION   AX779799
VERSION     AX779799.1  GI:32696793
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.,
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 1956 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE);
            Ludwig-Maximilian-Universitaet Muenchen (DE); Haferlach, Torsten,
            PD Dr. Dr. (DE); Schoch, Claudia (DE); Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1..1172
                     /organism="Homo sapiens"
```

```
Qy       421 GACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGGATTTGTAAAAATGCC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CTCTGGGGACCAAAAATCGAAAAAATGCATCAAAAAATGCCAATAAAAGAAAAAGTAAAAGT 540

Qy       481 CTCTGGGGACCAAAAATCGAAAAAATGCATCAAAAAATGCCAATAAAAGAAAAAGT 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TAACTTTTTGGTTTGATGTACACATATTCAAAAAAGTACATTAATAATGTCACAGTAA 600

Qy       541 TAACTTTTTTGGTTTGATGTACACATATTCAAAAAAGTACATTAATAATGTCACAGTAA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TATGTAAAGCTAAATACTTCCTCCAAAGATCATTATCTTTATTGATTGATTAGGACACTGAGGA 660

Qy       601 TATGTAAAGCTAAATACTTCCTCTCCAAAGATCATTATCTTTATTGATTGATTAGGACACTGAGGA 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 TTTTAACATTGTGATATATTAATTTATAATTTAATTTACCAATCTCTTGATGACAGTCTTATT 720

Qy       661 TTTTAACATTGTGATATATTAATTTATAATTTAATTTACCAATCTCTTGATGACAGTCTTATT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 TCTTTATATAGGTCAGTCTTGCAAGTACCATTTTATAGACAGCGTGTGAAATTTAAGTGAA 780

Qy       721 TCTTTATATAGGTCAGTCTTGCAAGTACCATTTTATAGACAGCGTGTGAAATTTAAGTGAA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ATGTTCTTTGTAAACATTTGTACTATTTTAAAATGAATAAGACCTTATGAAGTATGCTAT 840

Qy       781 ATGTTCTTTGTAAACATTTGTACTATTTTAAAATGAATAAGACCTTATGAAGTATGCTAT 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CTGTAGGCTGAAATTATAGGTACATCGTTTTCACTATGATATTAACAGAACGGTG-AA 899

Qy       841 CTGTAGGCTGAAATTATAGGTACATCGTTTTCACTATGATATTAACAGAACGGTGAA 900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       900 TGACTTAAAGTTCATTTTTTTCTGTATAGATACTTTATCATGTTTTCATGATTTTAGGA 959

Qy       901 TGACTTAAAGTTCATTTTTTTCTGTATAGATACTTTATCATGTTTTCATGATTTTAGGA 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       960 ATTACTGCTTTGTTGGATATTGTGAAACTAAAAGTTTATAGGGTTGTACTTTTAATTC 1019

Qy       961 ATTACTGCTTTGTTGGATATTGTGAAACTAAAAGTGTTTATAGGGTTGTACTTTTAATTC 1020
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1020 TTGGCATGTTGCCTCTATGTCCCATTTAAAATAAAATACATTTCTCATTAACTTTAGATGG 1079

Qy      1021 TTGGCATGTTGTGCCTCTATGTCCCATTTAAAATAAAATACATTTCTCATTAACTTTAGATGG 1080
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1080 GAAATAAGGTTGTATGTTGATGGATGATGAATTTGGCATGATGAATTTGGCAATAAAAGG 1139

Qy      1081 GAAATAAGGTTGTATGTTGATGGATGATGAATTTGGCATGATGAATTTGGCAATAAAAGG 1140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1140 CTGAAAAAGTTGTAAAA 1156

Qy      1141 CTGAAAAAGTTGTAAAA 1157
```

ORIGIN
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"

Query Match   99.0%; Score 1145; DB 6; Length 1172;
Best Local Similarity   99.9%; Pred. No. 5.8e-195;
Matches 1156; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

```
Qy     1 CTTTCCGGGAGACTGGAGTCGAAGGCCGTGAGTATTTCTTAAGCCAGTGTTTAGAGAGTA    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTTTCCGGGAGACTGGAGTCGAAGGCCGTGAGTATTTCTTAAGCCAGTGTTTAGAGAGTA    60

Qy    61 TGTGAGGCAAGAGTACCTTATAGAACCCCGAGGAGGGTGAGGAGCAGAGCCCATAATG   120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TGTGAGGCAAGAGTACCTTATAGAACCCCGAGGAGGGTGAGGAGCAGAGCCCATAATG   120

Qy   121 GCAGGTGAAGAAGAAATTAATGAAGACTATCCAGTAGTAGAAATTCACGAGTATTTGTCACGCGTTT   180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCAGGTGAAGAAGAAATTAATGAAGACTATCCAGTAGTAGAAATTCACGAGTATTTGTCAGCGTTT   180

Qy   181 GAGAATTCCATTGGTGCTCTGGATGAGATGCTGAAGACCATGATGTCTGTTCTTAGAAAT   240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GAGAATTCCATTGGTGCTCTGGATGAGATGCTGAAGACCATGATGTCTGTTCTTAGAAAT   240

Qy   241 GAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCATAC   300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCATAC   300

Qy   301 ACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACAT   360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACAT   360

Qy   361 CCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTTAAGGAAATAACA   420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTTAAGGAAATAACA   420

Qy   421 GACAAGAAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGAGATTTGTAAAAATGCC   480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACAAGAAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGAGATTTGTAAAAATGCC   480

Qy   481 CTCTGGGAACCAAAAATCGAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAGTAAAAGT   540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CTCTGGGAACCAAAAATCGAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAGTAAAAGT   540

Qy   541 TAACTTTTTGGTTTTGATGTACACATATTCAAAAAAGTACATTAATATGTAATCACAGTAA   600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TAACTTTTTGGTTTTGATGTACACATATTCAAAAAAGTACATTAATATGTAATCACAGTAA   600

Qy   601 TATGTAAAGCTAAATACTTCCTCTCCAAAGATCATTATCTTTATTGATTAGCACTGAGGA   660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TATGTAAAGCTAAATACTTCCTCTCCAAAGATCATTATCTTTATTGATTAGCACTGAGGA   660

Qy   661 TTTTAACATTGTGATATATATTTATAATTTACCATCTCTTGATGAGACTCTTATT   720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TTTTAACATTGTGATATATATTTATAATTTACCATCTCTTGATGAGACTCTTATT   720

Qy   721 TCTTTATATAGGTCAGTCTTGCAAGTACCAGCTGTGTGAAATTTAAGTGAA   780
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TCTTTATATAGGTCAGTCTTGCAAGTACCAGCTGTGTGAAATTTAAGTGAA   780

Qy   781 ATGTTCTTTGTAAAACATTTGTACTATTTTTAAAATGAATAATGAAGTAGTAGTATGCTAT   840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ATGTTCTTTGTAAAACATTTGTACTATTTTTAAAATGAATAATGAAGTAGTAGTATGCTAT   840

Qy   841 CTGTAAGGCTGAAATTATAGGTACATCTGTTTTCACTATATGATATTAAGAAAGGGTG-AA   899
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   841 CTGTAAGGCTGAAATTATAGGTACATCTGTTTTCACTATATGATATTAAGAAAGGGTGAAA   900

Qy   900 TGACTTAAAATGTTCATTTTTCTGTATAGATACTTTATCATGTTTTCATGATTTTAGGA   959
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TGACTTAAAATGTTCATTTTTCTGTATAGATACTTTATCATGTTTTCATGATTTTAGGA   960

Qy   960 ATTACTGCTTTGTTGTTGATATTCAAAGTGTGAAACTAAAAGTTATGGTTGTTGTACTTTAATTC  1019
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 ATTACTGCTTTGTTGTTGATATTCAAAGTGTGAAACTAAAAGTTATGGTTGTTGTACTTTAATTC  1019

Qy  1020 TTGGCATGTTGCCCTCTTATGTGCCCATTTAAAATAAAATACATTCATTACTTAGATGG  1079
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1020 TTGGCATGTTGCCCTCTTATGTGCCCATTTAAAATAAAATACATTCATTACTTAGATGG  1079

Qy  1080 GAAATAAGGTTGTATGTGATGAAATTTGATGACATTTAAAATAAAATACATTCATTAAGG  1139
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1080 GAAATAAGGTTGTATGTGATGAAATTTGATGACATTTAAAATAAAATACATTCATTAAGG  1139

Qy  1140 CTGAAAATGTTGTAAA  1156
         ||||||||||||||||
Db  1140 CTGAAAATGTTGTAAA  1156
```

RESULT 6
HSC1DPROT

LOCUS   HSC1DPROT   1172 bp   mRNA   linear   PRI 20-FEB-1998
DEFINITION H.sapiens mRNA for C1D protein.
ACCESSION   X95592
VERSION   X95592.1 GI:1185118
KEYWORDS   C1D gene; C1D protein; DNA-binding protein.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nehls,P., Keck,T., Greferath,R., Spiess,E., Glaser,T.,
            Rothbarth,K., Stammer,H. and Werner,D.
  TITLE   cDNA cloning, recombinant expression and characterization of
            polypeptides with exceptional DNA affinity
  JOURNAL   Nucleic Acids Res. 26 (5), 1160-1166 (1998)
  MEDLINE   98136194
   PUBMED   9469821
REFERENCE   2 (bases 1 to 1172)
  AUTHORS   Werner,D.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-FEB-1996) D. Werner, Dt. Krebsforschungszentrum,
            Biochemistry of the Cell, Im Neuenheimer Feld 280-0225, D-69120
            Heidelberg, FRG
FEATURES   Location/Qualifiers
     source   1..1172
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene   1..1172
                     /gene="C1D"
     CDS   118..543
                     /gene="C1D"
                     /note="DNA-binding protein"
                     /codon_start=1
                     /product="C1D protein"
                     /protein_id="CAA64845.1"
                     /db_xref="GI:1185119"
                     /db_xref="GOA:Q13901"
                     /db_xref="TrEMBL:Q13901"
                     /translation="MAGBEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQBLERIRVYMNRVKEITD
                     KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"
     polyA_signal   1050..1055
                     /gene="C1D"

ORIGIN

Query Match   99.0%; Score 1145; DB 9; Length 1172;
Best Local Similarity   99.9%; Pred. No. 5.8e-195;
Matches 1156; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

```
Qy     1 CTTTCCGGGAGACTGGAGTCGAAGGCCGTGAGTATTTCTTAAGCCAGTGTTTAGAGAGTA    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 7
BC009589
```

Qy   1140 CTGAAAATGTTGTAAAA 1156
           ||||||||||||||||||
Db   1141 CTGAAAAATGTTGTAAAA 1157

LOCUS   BC009589   1186 bp   mRNA   linear   PRI 29-JUN-2004
DEFINITION Homo sapiens nuclear DNA-binding protein, transcript variant 2,
            mRNA (cDNA clone MGC:14659 IMAGE:4102498), complete cds.
ACCESSION   BC009589
VERSION   BC009589.1 GI:16307025
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1186)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2 (bases 1 to 1186)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson,M., Schmutz,J., Grimwood,J., Rodriguez,A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 21 Row: b Column: 9
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 2784372.
FEATURES   Location/Qualifiers
     source   1..1186
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:14659 IMAGE:4102498"
                     /tissue_type="Bone marrow, chronic myelogenous leukemia"

Qy   1 CTTTCCGGGAGACTGGAGTTCGAAGGCCGTTGAGTGTGTTTAGAGAGTA 60
Db

Qy   61 TGTGAGGCAAGAGTAGCCTAGAGAACCCCGGAGGAGGAGGGGTGAGCAAGAGGCTGGCCATAATG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGTGAGGCAAGAGTAGCCTATAGAGAACCCCGGAGGAGGAGGGGTGAGCAAGAGGCTGGCCATAATG 120

Qy   121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTGTCAGCGGTTT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTGTCAGCGGTTT 180

Qy   181 GAGAATTCCATTGGTGTCGTGTGAGATGACGATGCTGAAGACCATGAGTTCTGTTTCTAGAAAT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GAGAATTCCATTGGTGTCGTGCTGTGGATGAGAGATGCTGAAGACCATGAGTTCTGTTTCTAGAAAT 240

Qy   241 GAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCATAC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCATAC 300

Qy   301 ACATTAAAATTCAAATGTTTTTGGGTTTAATTGGCCAACCCAAGGAGTTAATCCTTAGGAACAT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACATTAAAATTCAAATGTTTTTGGGTTTAATTGGCAACCCAAGGAGTTAATCCTTAGGAACAT 360

Qy   361 CCAGTAAAAACAGGAATTGGAAAGAATCAGTATATATATGAGACAGAGTCAAGGAAATAACA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCAGTAAAAACAGGAATTGGAAAGAATCAGTATATATATGAGACAGAGTCAAGGAAATAACA 420

Qy   421 GACAAGAAAAAGGCTGGCAAGCTGGACAGGGTCCAGCCTTCAGATTTGTTAGCAGAGGA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACAAGAAAAAGGCTGGCAAGCTGGACAGGGTCCAGCCTTCAGATTTGTTAGCAGAGGA 480

Qy   481 CTCTCGGGAACCAAAATCGAAAAATGCATCAAAGTTGCCAATAAGTTGCCAATAAGAAAGTAAAAGT 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CTCTCGGGAACCAAAATCGAAAAATCGAAAAAGTAGTTGCCAATAAGAGAAAAAGTAAAAGT 540

Qy   541 TAACTTTTGGTTTTGATGTACACATATTCAAAAGTACACATTAAATAATCACAGTAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TAACTTTTGGTTTTGATGTACACATATTCAAAAGTACACATTAAATAATCACAGTAA 600

Qy   601 TATGTAAAGCTAAATACTTCCTCTCCAAAGATCATATCTTTATTGATTAGCACTGAGGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TATGTAAAGCTAAATACTTCCTCTCCAAAGATCATATCTTTATTGATTAGCACTGAGGA 660

Qy   661 TTTTAACAATTGTGATATATATATTATATTTATTAATTTACCATCTCTTGATGACTCTTATT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TTTTAACAATTGTGATATATATATTATATTTATTAATTTACCATCTCTTGATGAGACTCTTATT 720

Qy   721 TCTTTATATAGGTCAGTGTCTTGCAAGTGAGCAGCCTGTGAAATTTAAGTGAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TCTTTATATAGGTCAGTGTCTTGCAAGTGAGCAGCCTGTGAAATTTAAGTGAA 780

Qy   781 ATGTTCTTTGTAAACATTTGTACTATTTAATGAATAATGAACCTTATGAAGTATAGCTTAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ATGTTCTTTGTAAACATTTGTACTATTTAATGAATAATGAACCTTATGAAGTATAGCTTAT 840

Qy   841 CTGTAGGCTGAAATTATAGGTACACTGTTTTCACTATATGATATTAAGAAAGCCGTG-AA 899
          ||||||||||||||||||||||||||||||||||||||||||||| ||
Db   841 CTGTAGGCTGAAATTATAGGTACACTGTTTTCACTATATGATATTAAGAAAGCCGTGAAA 900

Qy   900 TGACTTAAAGTCTTCATTTTTTTCGTATAGATAACTTTATTCATGATTTTAGGA 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TGACTTAAAGTCTTCATTTTTTTCGTATAGATAACTTTATTCATGATTTTAGGA 960

Qy   960 ATTACTGCTTTTGTTTGATTATTCAAAAGTTGTAACTTTAATTC 1019
          ||||||||||||||||||||||||||||||||||||||||||||
Db   961 ATTACTGCTTTTGTTTGATATTCAAAGTGAAAACTAAAAGTTGTAACTTTAATTC 1020

Qy   1020 TTGGCATGTTGCCCTCTAGTCCCATTAAAAATAAAATACATTCTCATTAACTTTAGATGG 1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTGGCATGTTGCCTCTAGTCCCATTAAAAATAAAATACATTCTCATTAACTTTAGATGG 1080

Qy   1080 GAAATAAGGTTGTATGTTGATGGATGAATTTCGGCATGATGACTGTACTCTCAATAAAGG 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GAAATAAAGGTTGTATGTTGATGGATGAATTTTGGCATGATGACTGTACTCTCAATAAAGG 1140

ORIGIN

```
gene          /clone_lib="NIH_MGC_54"
              /lab_host="DH10B"
              /note="Vector: pDNR-LIB"
              1..1186
              /gene="CID"
              /note="synonyms: MGC12261, MGC14659, SUNCOR"
              /product="nuclear DNA-binding protein"
              /db_xref="Locusid:10438"

CDS           /db_xref="AAH09589.1"
              50..475
              /gene="CID"
              /codon_start=1
              /product="nuclear DNA-binding protein"
              /protein_id="AAH09589.1"
              /db_xref="GI:16307026"
              /db_xref="Locusid:10438"
              /db_xref="MIM:606997"
              /translation="MAGEINEDYPVBIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
              QKLDPLEQAKVDLVSAYTLNSMFWVYLLATQGVNPKEHPVKQELERIRVYMNRVKEITD
              KKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"
```

Query Match   80.2%;   Score 927.2;   DB 9;   Length 1186;
Best Local Similarity   92.3%;   Pred. No. 5e-156;
Matches 1048; Conservative 0; Mismatches 18; Indels 70; Gaps 4;

```
Qy     91  GGAGGGTGAGGGAGGAGCTGGCCATTAATGGCCAGGTGAAGAAATTAATGAAGACTATCCA   150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     23  GGAGTCGGAAGGGCCCTGAGTCAGCCATTAATGGCAGGTGAAGAAATTAATGAAGACTATCCA   82

Qy    151  GTAGAAATTCACGAGTATTTGTCACGGTTTGAGAATTCATTGGTGCTGTGGATGAGATG   210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    83  GTAGAAATTCACGAGTATTTGTCACGGTTTGAGAATTCCATTGGTGCTGTGGATGAGATG   142

Qy    211  CTGAAGAACCATGATGTCTCGTTTCTTCAGAAATGAGTTGTTCGCTGGAGTTGGATCCACTTGAA   270
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   143  CTGAAGAACCATGAGTCTCGTTTCTTCAGAAATGAGTTGTTGCAGAATAAATCCACTTGAA   202

Qy    271  CAAGCAAAAGTGGATTTGGTTTCTGCATACACATTAAATTCAATGTTTTGGGTTTATTTG   330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203  CAAGCAAAAGTGGATTTGGTTTCTGCATACACATTAAATTCAATGTTTTGGGTTTATTTG   262

Qy    331  GCAACCCAAGGAGTTAATCCTAAGGAACATCCCAGTAAAACAGGAGAATTGGGAAAGAATCAGA   390
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   263  GCAACCCAAGGAGTTAATCCTAAGGAACATCCAGTAAAAACACAGGAATTGGAAAGAATCAGA   322

Qy    391  GTATATAATGGAACACAGAGTCAAGGAAATAACAGACAAGAAAAAGGCTGGGCAAGCTGGACAGA   450
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   323  GTATATAATGGAACACAGAGTCAAGGAAATAACAGACAAGAAAAAGGCTGGGCAAGCTGGACAGA   382

Qy    451  GGTGCAGCTTCAAGATTGTAAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAATCGATCA   510
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   383  GGTGCAGCTTCAAGATTTGTAAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAATGCATCA   442

Qy    511  AAAGTTGCCAATAAAGGAAAAAGTAAAAGTTAACTTTTTTGGTTTGATGTACACATATTC   570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   443  AAAGTTGCCAATAAAGGAAAAAGTAAACTTTTTTGGTTTGATGTACACATATTC   502

Qy    571  AAAAAAGTAC-----------------------------   579
           ||||||
Db   503  AAAAAGTACATCTCCCCCCCCCCGCAAAAATAATTCGTGGGCAGGGCAAGGTT   562

Qy    580  ------------ATTAATATGTAAATCACAG---TAATATGTAAAGCTAAAATAC-TTCC   621
           |||||||||     |||||||||||||||    ||||
Db   563  TAAAATGTGTTTCTTATTAATATGTAAAATTCACAGTAAAATATGTAAAGCTAAAAATACTTTCC   622

Qy    622  TCTCCAAAGATCATTATATCTTTATTGAGCACTGAGGATTTTAACATTGTGATTATATTA   681
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   623  TCTCCAAAGATCATTATCTTTATTGATTAGCACTGAGGATTTTAACATTGTTGATTATATTA   682

Qy    682  TATATTTATTAATTTACCATTCTTGATGATGAGACTCTTATTTCTTTATATAGGTCAGTCTTG   741
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   683  TATATTTATTAATTTACCATTCTTGATGATGAGACTCTTATTTCTTTATATAGGTCAGTCTTG   742

Qy    742  CAAGTACCATTTTATAAGCAGCTGGAAATTTAAGATGAAAATGTTCTTTGTAAACATTTGT   801
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   743  CAAGTACCATTTTATAAGCAGCTGGAAATTTAAGATGAAAATGTTCTTTGTAAACATTTGT   802

Qy    802  ACTATTTAAATGAATAATGACCTTATGAAGTATGCTATCTGTAGGCTGAAAATTATAGGT   861
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   803  ACTATTTAAATGAATAATGACCTTATGAAGTATGCTATCTGTAGGCTGAAAATTATAGGT   862

Qy    862  ACATCTGTTTTCACTATATAGATATTAAGAAAGCGTG-AATGACTTAAAATGTCATTTTT   920
           ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db   803  ACATCTGTTTTCACTATATAGATATTAAGAAAGCGTGATATGTCTATCTGTAGGCTGAAAATTATAGGT   862

Qy    921  TCTGTATAGATACTTTATCAGTTTCATGATTTAAGGAATTACTCGTTTGTTGATATTC   980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   863  ACATCTGTTTTCACTATATAGATATTAAGAAAGCGTGATATGTCTATCTGTAGGCTGAAAATTATAGGT   922

Qy    981  AAAGTGTGAAACTAAAAGTTTATGGTGTACTTAATTCTTGGCATGTGCCTTCTATGTC   1040
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   923  TCTGTATAGATACTTTATCAGTTTCATGATTTAAGGAATTACTCGTTTGTTGATATTC   982

Qy   1041  CCATTTAAAATAAAATACATTCTCATTAACTTTAGATGGGGAAATAAGGTTGTTGTTGAT   1100
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    983  AAAGTGTGAAACTAAAAGTTTATGGTGTACTTAATTCTTGGCATGTGCCTTCTATGTC   1042

Qy   1101  GGATGAATTTTGGCATGATGACTGTACTCTCCAATAAAAGGCTGAAAAAGTTGTAAA   1156
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1043  CCATTTAAAATAAAATACATTCTCATTAACTTTAGATGGGGAAATAAGGTTGTTAGTTGAT   1102

Db   1103  GGATGAATTTTGGCATGATGACTGTACTCTCCAATAAAAGGCTGAAAAAGTTGTATAA   1158
```

RESULT 8

Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu,
Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie
Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 15 Row: a Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.

FEATURES   Location/Qualifiers
source   1..1168
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:9308 IMAGE:3906517"
                     /tissue_type="Uterus, leiomyosarcoma"
                     /clone_lib="NIH_MGC_71"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
gene   1..1168
                     /gene="C1D"
                     /note="synonyms: MGC12261, MGC14659, SUNCOR"
                     /db_xref="LocusID:10438"
                     /db_xref="MIM:606997"
CDS   49..474
                     /gene="C1D"
                     /codon_start=1
                     /product="nuclear DNA-binding protein"
                     /protein_id="AAH16284.1"
                     /db_xref="GI:16740845"
                     /db_xref="LocusID:10438"
                     /db_xref="MIM:606997"
                     /translation="MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKIMMSVSRNEIL
                     QKLDPEQAKVDLVSAYTLNSMFWYYLATQGVNPKEHPVKQELERIRVYMNRVKEITD
                     KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"

ORIGIN

Best Local Similarity   80.1%;
Query Match   93.0%; Pred. No. 8.2e-156; Length 1168;
Matches 1040; Conservative   0; Mismatches   10; Indels   68; Gaps   4;

Qy   107 AGCTGGGCCATAAATGGCAGGTGAAGAAATTAATGAAGAACTATACCAGTAGAAATTCACGAGT 166
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   38 AGTCAGGCCATAATGGCAGGTGAAGAAAATAATGAAGAACTATCCAGTAGAAATTCACGAGT 97

Qy   167 ATTTGTCAGGCGTTTGAGAATTCCATTGGTTGCTGTGGAGATCGAGAGACCATGAGTGT 226
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   98 ATTTGTCAGGCGTTTGAGAATTCCATTGGTTGCTGTGGAGATCGAGAGACCATGAGTGT 157

Qy   227 CTGTTTTCTAGAAATCAGTTGTTGCAGAAGTTGGAATCACCAGCAAAAGTCGGATT 286
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   158 CTGTTTTCTAGAAATGAGTTGTTGCAGAAGTTGGAATCACCAGCAAAAGTCGGATT 217

Qy   287 TGGGTTTCTGCATACACACATTAAATTCAAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTA 346
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   218 TGGGTTTCTGCATACACACATTAAATTCAAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTA 277

Qy   347 ATCCTAAGGAAACATCCAGTAAAACAGGAAATTCAGAGTATATATGAACAGAG 406
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   278 ATCCTAAGGAAACATCCAGTAAAAACAGGAAATTCAGAGTATATATGAACAGAG 337

Qy   407 TCAAGGAAATAACAGACAAGAAAAAAGGCTGGCAAGCTGGACGAGAGGTTGCAGCTTCAAGAT 466
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   338 TCAAGGAAATAACAGACAAGAAAAAAGGCTGGCAAGCTGGACGAGAGGTTGCAGCTTCAAGAT 397

Qy   467 TTGTAAAAAATGCCCTCTTCGGGGAACCAAAATGAAAAATGCATCAAAAGTTGGCCAATAAAG 526
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   398 TTGTAAAAAATGCCCTCTTCGGGGAACCAAAATCGAAAAATGCATCAAAAGTTGGCCAATAAAG 457

Qy   527 GAAAAAGTAAAAGTTAACTTTTTTGATGTACACATATTCAAAAAGTAC------ 579
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   458 GAAAAAGTAAAAGTTAACTTTTTTGATGTACACATATTCAAAAAGTACACATCTTCC 517

Qy   580 --ATTA 583

Db   518 CCCCCCCCCCCCCGCAAAAAAATAATTCTGTGGCCAGGGCAAGGTTTAAAATGTGTTTCTTATTA 577

Qy   584 ATATGTAATTCACA---GTAATAGTTAAAGTTAAAATAC-TTCCCTCTCCAAAGAACCATTATC 639
         ||||||||||| ||||||||||||||||||||
Db   578 ATATGTAAATTCACAGTAATAGTTAAAGTTAAAATATATATCTTGCAAGTACCAATTTTATAAG 637

Qy   640 CAGCTGTGAAAATTGTATTTTGTTTGTTAAACATTGTATATATTATTATTATTATTAATTTTACCA 699
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   638 TTTATTGATTTAGTGCACTGAGGATTTTAACATTGTATATATTATTATTATTATTAATTTTACCA 699

Qy   700 TCTCTTGATAGACTCTTATTTCTTTATATAGGTCAGTGTTCTTCGCAAGTACCACCATTTTATAAG 759
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   698 TCTCTTGATAGAGACTCTTATTTCTTTATATAGGTCAGTGTTCTTGCAAGTACCATTTTATAAG 759

Qy   760 CAGCTGTGAAAGTTTCTTGTATATATTGTTTGTACTATTTTATTATTTATTAAATGAATAA 819
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   758 CAGCTGTGAAAATTTAAGTGAAAGTTCTTGTACTATTTTATTATTTATTAAATGAATAA 817

Qy   820 TGACCTTATGAAGTAGTGCTATCTGTAGGCTGAAAAATTATTAGGTAGTACATCGTTTTCACTATA 879
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   818 TGACCTTATGAAGTATGCTATCTGTAGGCTGAAAATTATTAGGTAGTACATCGTTTTCACTATA 879

Qy   880 TGATATTAAGAAAGGCGTG-AATGACTTAAAATGTTCATTTTTTTTCGTATAGACTTTAT 938
         |||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   878 TGATATTAAGAAAGGCGTGAAATGACTTAAAATGTTCATTTTTTTCGTATAGATAACTTTAT 937

Qy   939 CATGTTTTCATGATTTTAGGAAATTACTGCTTTGTTGAATTCAAAGTGTGAAACTAAAAG 998
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   938 CATGTTTTCATGATTTTAGGAAATTACTGCTTTGTTGAATTCAAAGTGTGAAACTAAAAG 997

Qy   999 TTTATGGTTGTTGTACTTTAATTCTTGGCATGTTGCCTCTCTATGTCCCATTTAAAATTAAAATAC 1058
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   998 TTTATGGTTGTTGTACTTTAATTCTTGGCATGTTGCCTCTCTATGTCCCATTTAAAATTAAAATAC 1057

Qy 1059 ATTCTCCATTAACTTTAGATGGGGAAATAAGATTGTTGTTTGATGGATGGAATTTTGGCATGA 1118
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1058 ATTCTCATTAACTTTAGATGGGGAAATAAGATTGTTGTTTGATGGATGGAATTTTGGCATGA 1117

Qy 1119 TGACTGTACTCTCAATAAAGGCTGAAAATGTTGTAAAA 1156
         ||||||||||||||||||||||||||||||||||||||
Db 1118 TGACTGTACTCTCAATAAAGGCTGAAAATGTTGTATATA 1155

LOCUS   BC009584   1195 bp   mRNA   linear   PRI 29-JUN-2004

Query Match   79.8%;   Score 922.4;   DB 9;   Length 1195;

ORIGIN

REFERENCE   2   (bases 1 to 1195)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (29-JUN-2001) National Institutes of Health, Mammalian
   Gene Collection (MGC), Cancer Genomics Office, National Cancer
   Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
   USA
COMMENT   NIH-MGC Project URL: http://mgc.nci.nih.gov
   Contact: MGC help desk
   Email: cgapbs-r@mail.nih.gov
   Tissue Procurement: ATCC
   cDNA Library Preparation: CLONTECH Laboratories, Inc.
   cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
   DNA Sequencing by: Sequencing Group at the Stanford Human Genome
   Center, Stanford University School of Medicine, Stanford, CA 94305
   Web site: http://www-shgc.stanford.edu
   Contact: (Dickson, Mark) mcd@paxil.stanford.edu
   Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
   R. M.
   Clone distribution: MGC clone distribution information can be found
   through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
   Series: IRAL Plate: 21 Row: n Column: 3.

FEATURES   Location/Qualifiers
   source   1..1195
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="MGC:14647 IMAGE:4095552"
      /tissue_type="Bone marrow, chronic myelogenous leukemia"
      /clone_lib="NIH_MGC_54"
      /lab_host="DH10B"
      /note="Vector: pDNR-LIB"
   gene   1..1195
      /gene="C1D"
      /note="synonyms: MGC12261, MGC14659, SUNCOR"
      /db_xref="LocusID:10438"
      /db_xref="MIM:606997"
   CDS   56..481
      /gene="C1D"
      /codon_start=1
      /product="nuclear DNA-binding protein"
      /protein_id="AAH09584.1"
      /db_xref="GI:16307018"
      /db_xref="LocusID:10438"
      /db_xref="MIM:606997"
      /translation="MAGEEINEDYPVEIHEVLSAFENSIGAVDEMLKTMMSVSRNELL
      QKLDPLEQAAKVDLVSAYTLNSMFWVLATQGVNPKEHPVKQELERIRVYMRVKEITD
      KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"

REFERENCE   1   (bases 1 to 1195)
AUTHORS   Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
   Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
   Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
   Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
   Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
   Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
   Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
   Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
   McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
   Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
   Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
   Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodriguez,S.,
   Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
   Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
   Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
   Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
   Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE   Generation and initial analysis of more than 15,000 full-length
   human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED   12477932
REMARK

Best Local Similarity   92.8%;   Pred. No. 3.6e-155;
Matches 1039;   Conservative   0;   Mismatches   11;   Indels   70;   Gaps   4;

Qy   107 AGCTGGCCCATAATGCGACCAGGTGGAAGAAGAATTAANTGAAGACTATCCAGTTAGAAANTTCACGAGT 166
   |||||||||||||||||||||||| |||||||||| ||||||||||||| ||||||||| ||||||||||
Db   45 AGTCAGCCATAATGCGACCAGGTGGAAGAAGAATTAATGAAGACTATCCAGTTAGAAATTCACGAGT 104

Qy   167 ATTTGTCAGCGTTTGAGAATTCCATTGGTGCTGTGGATGAGAGTCTGAAGACCATGATGT 226
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   105 ATTTGTCAGCGTTTGAGAATTCCATTGGTGCTGTGGATGAGAGTCTGAAGACCATGATGT 164

Qy   227 CTGTTTCTAGAAATGACTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATT 286
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   165 CTGTTTCTAGAAATGACTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATT 224

Qy   287 TGGGTTTCTGCATACACATTAAATTCAATGTTTGGGTTTATTGGCAACCCAAGGAGTTA 346
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   225 TGGGTTTCTGCATACACATTAAATTCAATGTTTGGGTTTATTGGCAACCCAAGGAGTTA 284

Qy   347 ATCCTAAGGACACATCCAGTAAAACACGGAATTGGAAAGAAATCAGAGTATATATGAACAGAG 406
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   285 ATCCTAAGGACACATCCAGTAAAACACGGAATTGGAAAGAAATCAGAGTATATATGAACAGAG 344

Qy   407 TCAAGGAAATAACAGACAAGAAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGAT 466
   ||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db   345 TCAAGGAAATAACAGACAAGAAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGAT 404

Qy   467 TTGTAAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAATGCATCAAAAGTTGGCAATAAAG 526
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   405 TTGTAAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAATGCATCAAAAGTTGGCAATAAAG 464

Qy   527 GAAAAAGTAAAATGTAAACTTTTTGGTTTGATGAGTACACATATTCAAAAAGTAC 579
   |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   465 GAAAAAGTAAAAGTTAACTTTTTGGTTTGATGAGTACACATATTCAAAAAGTACACTCTTCC 524

Qy   580 ---------------------GTAATATGTAAAGCTAAATAC--TTCCCTCTCCAAAGATCATTA 637
          ||||||||||||||||| ||||||||
Db   525 CCCCCCCCCCCCCCCGCAAAAAATAAATTCGTGGCAGGGCAAGGTTTAAAATGGTTTCTTAT 584

Qy   582 TAATATGTAAAGCTAAATAC---GTAATATGTAAAGCTAAATAC-TTCCCTCTCCAAAGATCATTA 644
   ||||||||||||||||||||
Db   585 TAATAATGTAAATTCACAGTAAATATGTAAAGCTAAATACTTTCCTCCTCCAAAGATCATTA 644

Qy   638 TCTTTATTGATTAGCACTGAGGATTTTAACATTGTGATATATATATATTATATTATTAC 697
   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   645 TCTTTATTGATTAGCACTGAGGATTTTAACATTGTGATATATATATATTATATTATTAC 704

Qy   698 CATCTCTTGATGAGACTCTTATTTCTTATTATAGGTCAGTGTTGCAAGTCGACCATTTTATA 757
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   705 CATCTCTTGATGAGACTCTTATTTCTTATTATAGGTCAGTGTTGCAAGTCGACCATTTTATA 764

Qy   758 AGCAGCGTGCGAAAATTTAAGTGAAAANTGTTCTTGTAAACATTTGTACTATTTTTAAAATGAAT 817
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   765 AGCAGCGTGCGAAAATTTAAGTGAAAAGTTCTTGTAAACATTTGTACTATTTTTAAAATGAAT 824

Qy   818 AATGACCTTATGAAGTATGTCTATCCTGGTAGGCTGAAATTATAGGTACATCGTGTTTTCACTA 877
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   825 AATGACCTTATGAAGTATGTCTATCCTGGTAGGCTGAAATTATAGGTACATCGTGTTTTCACTA 884

Qy   878 TATGATATTAAGAAAGCGTG--AATGACTAAATGTTCATTTTTTTCTGTATAGATACTTT 936
   |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   885 TATGATATTAAGAAAGCGTGAAATGACTAAATGTTCATTTTTTTCTGTATAGATACTTT 944

Qy   937 ATCACATGTTTCATGATTTTAGGAATTACTGGCTTGATATTCAAAGTGTGAAAACTAAA 996
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   945 ATCACATGTTTCATGATTTTAGGAATTACTGGCTTGATATTCAAAGTGTGAAAACTAAA 1004

Qy   997 AGTTTATGGGTTGTACTTTAAATTCTTGGCATGGTGCCTCTATGTCCCATTAAAATAAAAAT 1056
   || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1005 ACTTTATGGGTTGTACTTTAAATTCTTGGCATGGTGCCTCTATGTCCCATTAAAATAAAAAT 1064

Qy   1057 ACAATCTCATTAACTTTAGATGGGAAATAAGGTTGTATTGTTGATGGATGAATTTTGGGCAT 1116
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1005 ACAATCTCATTAACTTTAGATGGGAAATAAGGTTGTATTGTTGATGGATGAATTTTGGGCAT 1064

```
RESULT 10
Qy    1065 ACAATTCTCATTAACTTTAGATGGGAATAAGGTTGTATGTTGATGGATGAATTTCGGCAT 1124
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1117 GATGGATGAATTTCGGCAT 1156

Qy    1117 GATGGATGAATTTCGGCAT 1156
                |||||||||||||||||||
Db    1125 GATGGACTGTACTCTCAATAAAGGCTGGAAAATGTTGTATAA 1164
```

LOCUS   BC005235   1199 bp   mRNA   linear   PRI 29-JUN-2004
DEFINITION Homo sapiens nuclear DNA-binding protein, mRNA (cDNA clone MGC:12261 IMAGE:3930648), complete cds.
ACCESSION   BC005235
VERSION   BC005235.1 GI:13528872
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 1199)
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED   12477932
REFERENCE   2 (bases 1 to 1199)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (27-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: Dickson, Mark
                              mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 16 Row: 1 Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 27894371.

FEATURES   Location/Qualifiers
     source   1..1199
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:12261 IMAGE:3930648"
     gene   1..1199
                     /gene="C1D"
                     /note="synonyms: MGC12261, MGC14659, SUNCOR"
                     /clone_lib="NIH_MGC_59"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
                     /db_xref="LocusID:10438"
                     /db_xref="MIM:606997"
     CDS   64..489
                     /gene="C1D"
                     /codon_start=1
                     /product="nuclear DNA-binding protein"
                     /protein_id="AAH05235.1"
                     /db_xref="GI:13528873"
                     /db_xref="LocusID:10438"
                     /db_xref="MIM:606997"
                     /translation="MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
                     QKLDPLEQARKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERIRYYMNRVKEITD
                     KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"

ORIGIN
     Query Match   79.8%; Score 919.8; DB 9; Length 1199;
     Best Local Similarity   92.6%;
     Matches 1038; Conservative   0; Mismatches 12; Indels 71; Gaps   4;

Qy   107 AGCTGGCCATAATAATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGT 166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   53 AGTCAGCCATAATGCCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGT 112

Qy   167 ATTTGTCAGCGTTGAGAATTCCATTGGTGCTGTGGATGAAGACGATGATGTGAAGACCATGATGT 226
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   113 ATTTGTCAGCGTTGAGAATTCCATTGGTGCTGTGGATGAAGACGATGATGTGAAGACCATGATGT 172

Qy   227 CTGTTTCTAGAAATGAGTTGTTGCAGAAGTTGGATCCACTTGAAACAAGCAAAAGTGGATT 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   173 CTGTTTCTAGAAATGAGTTGTTGCAGAAGTTGGATCCACTTGAAACAAGCAAAAGTGGATT 232

Qy   287 TGGTTTCTGCATACACATTAAAATTCAATGTTTTGGGTTTTTATTTGGGCAACCCAAGGAGTTA 346
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   233 TGGTTTCTGCATACACATTAAAATTCAATGTTTTGGGTTTTTATTTGGGCAACCCAAGGAGTTA 292

Qy   347 ATCCTAAGGAACATCCAGTAAAACAGGAAATTCAGAGTATATATGAACAGAG 406
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   293 ATCCTAAGGAACATCCAGTAAAACAGGAAATTCAGAGTATATATGAACAGAG 352

Qy   407 TCAAGGAAATAACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCACCTTCANAGAT 466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   353 TCAAGGAAATAACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGAT 412

Qy   467 TTGTAAAAAATGCCCTCTGGGAGCCAAAAATCGAAAAATGCATCAAAAAGTTGCCAATAAAG 526
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   413 TTGTAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAATGCATCAAAAAGTTGCCAATAAAG 472

Qy   527 GAAAAAGTAAAAGTAACTTTTTGGTTTTGATGTACACATATTCAAAAAGTAC------ 579
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   473 GAAAAAGCAAAAGTTAACTTTTTGGTTTTGATGTACACATATTCAAAAAGTACATCTTCC 532

Qy   580 --A 580
                                                                      |
Db   533 CCCCCCCCCCCCCCCCCCCGAAAAATAATTCTGTGGCAGGGCAAAGGTTTAAATGTGGTTTCTTTA 592

Qy   581 TTAAATATGTAATCACA---GTAATATGTAAAGCTAAAATAC-TTCCTCTCCAAAGATCATT 636
              ||||||||||||||||| ||||||||||||||||| ||||||||||||||||||
Db   593 TTAAATATGTAAATTCACAGTAAATATGTAAAGCTAAAATAC-TTCCTCTCCAAAGATCATT 652

Qy   637 ATCTTTATTTGATTAGCCACTGAGGATTTAACATTGTGATATTATTTTATTATTAATTTA 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   653 ATCTTTATTTGATTAGCCACTGAGGATTTAACATTGTGATATTATTTTATTATTAATTTA 712

Qy   697 CCATCTCTTGATGAGAGCTCTTATTTCTTTATATATAGGTCAGTCTTGCAAGTACCATTTAT 756
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   713 CCATCTCTTGATGAGAGCTCTTATTTCTTTATATATAGGTCAGTCTTGCAAGTACCATTTAT 772
```

```
Qy    757  AAGCAGCTGTGAAATTAAGTGAAAAGTTTCTTGTGTAAACATTGTACTATTTAAATGA  816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    773  AAGCAGCTGTGAAATTAAGTGAAAAGTTTCTTGTGTAAACATTGTACTATTTAAATGAA  832

Qy    817  TAATGACCTTATGAAGTATCCTATCTGTAGGCTGAAATTATAGGTGACATCTGTTTTCACT  876
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    833  TAATGAACCTTATGAAGTATCCTATCTGTAGGCTGAAATTATAGGTGACATCTGTTTTCACC  892

Qy    877  ATATGATAATTAAGAAAAAGCCTG-AATGACTTAAATGTTCATTTTTTTCTGTATAGATACTT  935
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    893  ATATGATAATTAAGAAAAAGCCTGAAATGACTTAAATGTTCATTTTTTTCTGTATAGATACTT  952

Qy    936  TATCAGTGTTTTCATGATTTTAGGAATTAACTACGCTTTGTTGATATTCAAAGTGTGAAACTAA  995
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    953  TATCAGTGTTTTCATGATTTTAGGAATTAACTACGCTTTGTTGATATTCAAAGTGTGAAACTAA  1012

Qy    996  AAGTTTATGTTGTACTTTTAATTCTTGGCATGTTGCCCTCTATGTTCCCCATTTAAAATAAAA  1055
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1013  AACTTTATGGTTGTACTTTAATTCTTGGCATGTTGCCCTCTATGTTCCCCATTTAAAATAAAA  1072

Qy   1056  TACATTCTCATTAACTTTAGATGGGAAAATAAAGGTTGTATGGTTGTATGGGATGAAAATTTTTGGCA  1115
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1073  TACATTCTCATTAACTTTAGATGGGAAAATAAAGGTTGTATGGTTGTATGGGATGAAATTTTTTGGCA  1132

Qy   1116  TGATGACTGTACTCTCAATAAAGGCTGAAAAATGTTGTAAAA  1156
           ||||||||||||||||||||||||||||||||||||||||||
Db   1133  TGATGACTGTACTCTCAATAAAGGCTGAAAAATGTTGCAAAA  1173
```

```
RESULT 11
AL354750/c

LOCUS       AL354750       186856 bp    DNA    linear    PRI 03-DEC-2001
DEFINITION  Human DNA sequence from clone RP11-19501 on chromosome 10 Contains
            two novel (pseudo)genes, the 5' end of a gene for a novel protein
            and 2 CpG islands, complete sequence.
ACCESSION   AL354750
VERSION     AL354750.12  GI:10715924
KEYWORDS    HTG; CpG island; transport protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186856)
  AUTHORS   Lawlor,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 6, 2000 this sequence version replaced gi:10129439.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.   Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            This sequence is the entire insert of clone RP11-19501 The true
            left end of clone RP11-453G10 is at 166973 in this sequence. The
            true right end of clone RP11-135A24 is at 6540 in this sequence.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30.); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
```

```
FEATURES             Location/Qualifiers
     source          1..186856
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-19501"
                     /clone_lib="RPCI-11.1"
     misc_feature    3..448
                     /note="match: GSS: Em:AQ379747"
     misc_feature    complement(6088..6517)
                     /note="match: GSS: Em:AQ520457"
     misc_feature    7..461
                     /note="match: GSS: Em:AQ151771"
     misc_feature    complement(6171..6534)
                     /note="match: GSS: Em:AQ766212"
     misc_feature    7..448
                     /note="match: GSS: Em:AQ147832"
     misc_feature    6545..7024
                     /note="match: GSS: Em:AQ561548"
     misc_feature    115..429
                     /note="match: GSS: Em:AQ436650"
     misc_feature    6588..7012
                     /note="match: GSS: Em:AQ324035"
     repeat_region   2182..2217
                     /note="match: GSS: Em:AQ413399"
     repeat_region   10015..10842
                     /note="L1PA14 repeat: matches 5314..6149 of consensus"
     repeat_region   3383..3408
                     /note="18 copies 2 mer to 91% conserved"
     repeat_region   11148..12383
                     /note="L1PA14 repeat: matches 4086..5314 of consensus"
     repeat_region   5428..5511
                     /note="42 copies 2 mer ta 82% conserved"
     repeat_region   12901..13700
                     /note="L1PA14 repeat: matches 8..763 of consensus"
     misc_feature    complement(5845..6532)
                     /note="match: GSS: Em:AQ379747"
     misc_feature    17708..18165
                     /note="LTR41 repeat: matches"
     misc_feature    18643..19068
                     /note="match: STS: Em:G52882"
     misc_feature    25943..26064
                     /note="match: GSS: Em:AQ202290"
     repeat_region   32371..32545
                     /note="L1PB3 repeat: matches 6026..6147 of consensus"
     repeat_region   36682..38127
                     /note="L1M3c repeat: matches 1140..1315 of consensus"
     repeat_region   38145..42839
                     /note="L1MA8 repeat: matches 4783..6291 of consensus"
     misc_feature    43906..44653
                     /note="L1PA7 repeat: matches 1433..6141 of consensus"
     repeat_region   44803..46255
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    49657..50116
                     /note="L1PA7 repeat: matches 9..1460 of consensus"
                     /note="CpG island"
                     /evidence=non_experimental
     gene            complement(54351..55469)
                     /gene="bA19501.1"
                     /gene="bA19501.1"
     mRNA            complement(54351..>55469)
                     /gene="bA19501.1"
                     /product="bA19501.1 (novel protein similar to C1D
                     DNA-binding protein)"
                     /note="match: cDNAs: Em:X95592 Em:X95591
                     match: ESTs: Em:AA831747 Em:T83118 Em:AA352554 Em:AI004680
```

```
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP11-19501 is from
the library RPCI-11.1 constructed by the group of Pieter de Jong.
For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
```

```
polyA_site      Em:AA442782 Em:AA305785 Em:AI804147 Em:AI809071
                Em:AI800846 Em:AI351827 Em:AA835937 Em:N40848
                Em:AI553826 Em:AA458588 Em:AI831757 Em:N42131 Em:AI935349
                Em:N71224 Em:AW118590 Em:AW169841 Em:AAA31129 Em:M87910
                Em:AI243111 Em:AA424253 Em:AW024957 Em:AI699498
                Em:AI989469 Em:AI310127 Em:AA418433 Em:AI123353
                Em:AI800532 Em:AI507721 Em:AI460755 Em:R68158 Em:AA861428
                Em:AI274315 Em:AA424184 Em:AI218433 Em:AI027628
                Em:AI301876 Em:AA476674 Em:AI375251 Em:AI027628
                Em:AI476155 Em:AA476674 Em:AI816911 Em:AA861428
                Em:AW340622 Em:AI435863 Em:AI435192 Em:AW104208
                            Em:AI379715 Em:AI671948 Em:W01341 Em:AW592440"
                /gene="bAl950i.1"
                /evidence=not experimental
polyA_signal    complement(54383..54388)
                /gene="bAl950i.1"
misc_feature    54386..54689
                /gene="bAl950i.1"
                /note="match: STS: Em:G23952"
misc_feature    complement(54999..55406)
                /gene="bAl950i.1"
                /note="match: GSS: Em:AQ461100"
CDS             complement(55051..55386)
                /gene="bAl950i.1"
                /note="possible pseudogene
                match: proteins: Tr:Q13901"
                /codon_start=1
                /evidence=not experimental
                /product="bAl950i.1 (novel protein)"
                /protein_id="CAD13204.1"
                DNA-binding protein)"
                /db_xref="GI:17384063"
                /translation="MLKTTMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQ
                GVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKNALWEPKPKNASK
                VANKGKSKS"
misc_feature    55412..55487
                /note="match: GSS: Em:AQ463761 Em:AQ505503"
repeat_region   61512..61611
                /note="50 copies 2 mer aa 61% conserved"
repeat_region   68837..68982
                /note="LiPA5 repeat: matches 5997..6143 of consensus"
repeat_region   71617..71998
                /note="LiP83 repeat: matches 5769..6150 of consensus"
repeat_region   72363..72506
                /note="72 copies 2 mer ta 77% conserved"
misc_feature    78102..78408
                /note="match: GSS: Em:AQ628574"
misc_feature    complement(78524..78950)
                /note="match: GSS: Em:AQ401519"
repeat_region   78659..78758
                /note="LiM3b repeat: matches 102..199 of consensus"
repeat_region   79868..80242
                /note="LTR38 repeat: matches 129..503 of consensus"
repeat_region   80250..84310
                /note="LiPA15 repeat: matches 2098..6155 of consensus"
gene            84306..84837
                /note="LiPA15 repeat: matches 1407..1939 of consensus"
                join(87084..87121,87983..88037,109294..109356,
                111406..111627,115595..115629,118165..118302,
                126770..127043,127968..128040,177990..178100)
                /gene="bA479G22.1"
mRNA            join(87084..87121,87983..88037,109294..109356,
                111406..111627,115595..115629,118165..118302,
                126770..127043,127968..128040,177990..>178100)
                /gene="bA479G22.1"
                /product="bA479G22.1"
                /note="isoform 2
                continues as bA453G10.1 in Em:AL356053
                match: cDNAs: Em:AK023093
                match: ESTs: Em:AA421130"
                /evidence=not_experimental
mRNA            join(87084..87121,87983..88037,109294..109356,
                111406..111627,126770..127043,127968..128040,
                177990..>178100)
```

```
misc_feature    /gene="bA479G22.1"
                /product="bA19501.3 (novel protein)"
                /note="isoform 1
                continues as bA453G10.1 in Em:AL356053
                match: cDNAs: Em:AK023093
                match: ESTs: Em:AA421130"
                /evidence=not_experimental
                complement(87055..87660)
                /note="match: GSS: Em:AQ474468"
repeat_region   88133..88174
                complement(88502..88909)
misc_feature    /note="21 copies 2 mer tg 83% conserved"
                /note="match: GSS: Em:AQ129993"

Query Match     77.9%;    Score 900.4;  DB 9;  Length 186856;
Best Local Similarity  91.2%;   Pred. No. 1.2e-151;
Matches 1040;  Conservative  0;  Mismatches  26;  Indels  74;  Gaps  5;
```

```
Qy    91 GGAGGGTGAGGAGGACAGAGGCTGGCCATAATGGCAGGTGAAGAAATTAATAGAACTATCCA 150
         ||||| |||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db 55503 GGAGTCGAAGGCGGTGAGTCAGCCATAACCGGCAGGTGAAGAAATTAATGAAGACTATCCA 55444

Qy   151 GTAGAAATTCACGAGTATTTGTTCAGCGTTTGAGAATTCCATTGGTGGCTGGATGGAGATG 210
         ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db 55443 GTAGAAATTCACGAGTATTTGTTCAGCGTTTGAGAATTCCATTGGTGGCTGGATGGAGATG 55384

Qy   211 CTGAGAGACCATGATGTCTCGTTTCTTCCATACACATTAAATTCAATGTTTTCGCACTTGAA 270
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 55383 CTGAAGAGACCACGATCGTCTCGTTTCTTCGAAATGAGTTGTTGCAGAAATGAGTCCACTTGAA 55324

Qy   271 CAAGCAAAAAGTGGATTTGGTTCTTCGCATACAATTAAATTCAATGTTTTGGGTTTATTTG 330
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 55323 CAAGCAAAAAGTGGATTTGGTTCTTCGCATACAATTAAATTCAATGTTTTGGGTTTATTTG 55264

Qy   331 GCAACCCAAGGAGTTAATCCTAAGGAACATCCAGTAAAAACACAGGAAATTGGAAAGAATCAGA 390
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 55263 GCAACCCAAGGAGTTAATCCTAAGGAACATCCAGTAAAAACACAGGAAATTGGAAAGAATCAGA 55204

Qy   391 GTATATATGGAACAGAGTCAAGGAAATAACAGACAAGAGAAAAGGCTGGCAAGCTGGACAGA 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 55203 GTATATGGAACAGAGTCAAGGAAATAACAGACAAGAGAAAAGGCTGGCAAGCTGGACAGA 55144

Qy   451 GGTGCAGCTTCAAGATTTGTTAAAAAAATCGCCCTCTGGGAACCAAAAATCGAAAAATGCATCA 510
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 55143 GGTGCAGCTTCAAGATTTGTTAAAAAAATCGCCCTCTGGGAACCAAAAATGCATCA 55084

Qy   511 AAAGTTGCCAATAAAAGGAAAAAAGTAAAAGTTAACTTTTTTGGTTTTGTGTACACATATTC 570
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 55083 AAAGTTGCCAATAAAAGGAAAAAAGTAAAAGTTAACTTTTTTGGTTTTGTGTACACATATTC 55024

Qy   571 AAAAAAGTAC---------------------------------------------------- 570
         ||||||||||
Db 55023 AAAAAAGTACATTTCCCCGCCCCCCCCCACGAATAATTCTGTGGCAGGACAAGG 54964

Qy   580 ------------ATTAATATGTGAATCACAG---TAATATGTAAAAGCTAAAAATAC-TT 619
                     ||||||||||||||||||||    ||||||||||||||||||||||||||
Db 54963 TTTAAAAGTGTTCTTTATTAATAGTAAAATTCACAGTAAAATAATGTAAAGCTAAAAATACTTT 54904

Qy   620 CCTCTCCAAAGATCATTATATCTTTATTGATTAGCACTGAGCAGATTTAACATTGTGATATAT 679
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 54903 CCTCTCCAAAGATCATTATATCTTTATTGATTAGCACTGAGCAGATTTAACATTGTGATATAT 54844

Qy   680 TATATATATTTATATAATTTTACCATCTC--TTGATGAGACTCTTATTTCTTTATATAGGTCAGT 737
         |||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db 54843 TATATATATTTATATAATTTTACCATCTCTTTGATGAGACTCTTATTTCTTTATATAGGTCAGT 54784

Qy   738 CTTGCAAGTACCATTTATTAAGCACGCTGTGAAAATTAAGTAGAAAGTTCTTTGTAAAACAT 797
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 54783 CTTGCAAGTACCATTTATTAAGCACGCTGTGAAAATTAAGTAGAAAGTTCTTTGTAAAACAT 54724

Qy   798 TTGTACTATTTAAATGAATAATGACCTTATGAAGTATGCTATCTGTAGGCTGAAAATTAT 857
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 54723 TTGTACTATTTAAATGAATAATGACCTTATGAAGTATGCTATCTGTAGGCTGAAAATTAT 54664
```

```
Qy    858  AGGTACATCTGTTTTCACTATATGATATTAAGAAAGGCGTG-AATGACTTAAATGTTCATT  916
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  54663  AGGCACATCTGTTTTCACTATATGATATTAAGAAAGACGAAAATGACTTAAATGTTCATT  54604

Qy    917  TTTTTCTGTATAGATACTTTATCATGTTTCATGATTTTAGGAATTACTCGCTTTGTTGAT  976
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54603  TTTTTCTGTATAGATACTTTATCATGTTTCATGATTTTAGGAATTACTCGCTTTGTTGAT  54544

Qy    977  ATTCAAAAGTGTGAAACTAAAAGTTTATAGTTTGTACTTTAATTCTTTGGCATTGTTGCCTCTA  1036
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54543  ATTCAAAAGTGTAAAACTAAAACTTTATAGGTTGTACTTTAATTCTTTGGCATTGTTGCCTCTA  54484

Qy   1037  TGTCCCATTTAAAATAAAAATACATTCTCATTAACTTAGATGGGAAATAAGGTTGTATGT  1096
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  54483  TGTCCCATTTAAAATAAAAATACATTCTCATTAACTTAGATGGGAAATAAGGTTGTATGT  54424

Qy   1097  TGATGGATGAATTTTGGCATGATGACTGACTGTACTCTCCAATAAAAGGCTGAAAATGTTGTAAAA  1156
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db  54423  TGATGGATGAATTTTGGCATGATGACTGTACTCTCCAATAAAAGGCTGAAAATGTTGTATAA  54364
```

```
RESULT 12
LOCUS       AC068741
ACC068741/c
DEFINITION  Homo sapiens clone RP11-777E12, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   AC068741                195426 bp    DNA     linear    HTG 22-NOV-2000
VERSION     AC068741.3  GI:11276256
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 195426)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-777E12
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 195426)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavky,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Colmore,J., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karataas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,J., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,F., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (08-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 22, 2000 this sequence version replaced gi:9369524.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp-genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
```

```
            Center project name: L6655
            Center clone name: 777_E_12
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 30% of reads
            Sequencing vector: Plasmid; n/a; 70% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 191262 bases at least Q40
            Consensus quality: 193601 bases at least Q30
            Consensus quality: 194390 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 194726; sum-of-contigs
            Quality coverage: 6.9 in Q20 bases; agarose-fp
            Quality coverage: 6.5 in Q20 b.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *      1       731: contig of 731 bp in length
            *    732       831: gap of 100 bp
            *    832      3248: contig of 2417 bp in length
            *   3249      3348: gap of 100 bp
            *   3349      8620: contig of 5272 bp in length
            *   8621      8720: gap of 100 bp
            *   8721     17691: contig of 8971 bp in length
            *  17692     17791: gap of 100 bp
            *  17792     25774: contig of 7983 bp in length
            *  25775     25874: gap of 100 bp
            *  25875     66476: contig of 40602 bp in length
            *  66477     66576: gap of 100 bp
            *  66577    128195: contig of 61619 bp in length
            *  128196    128295: gap of 100 bp
            *  128296    195426: contig of 67131 bp in length.
FEATURES             Location/Qualifiers
     source          1..195426
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-777E12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..731
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    832..3248
                     /note="assembly_fragment"
     misc_feature    3349..8620
                     /note="assembly_fragment"
     misc_feature    8721..17691
                     /note="assembly_fragment"
     misc_feature    17792..25774
                     /note="assembly_fragment"
     misc_feature    25875..66476
                     /note="assembly_fragment"
     misc_feature    66577..128195
                     /note="assembly_fragment"
     misc_feature    128296..195426
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN
```

```
Query Match          77.9%;   Score 900.4; DB 2;  Length 195426;
Best Local Similarity 91.2%;   Pred. No. 1.1e-151;
Matches 1040; Conservative   0; Mismatches  26; Indels  74; Gaps   5;

Qy     91  GGAGGGTGAGGAGCAGAGCTGGCCATAATGGCCAGGTGAAGAAAATTAATGAAGAGACTATCCA  150
           ||||| |||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db  98376  GGAGTCGAAGGCCGTGAGTCAGCCATAACGGCCAGGTGAAGAAAAATTAATGAAGAGACTATCCA  98317
```

```
Qy    151 GTAGAAATTCACGAGTATTTGTCAGCCGTTTGACGAATTCCATTGGTGTGTGGATGAGATG 210
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  98316 GTAGAAATTCACGAGTATTTGTCAGCCGTTTGACGAATTCCATTGGTGCTGTGGATGAGATG 98257

Qy    211 CTGAAGACCATGATGTCTGTTTCTAGAAATGAGTTGTTGCAGAAGTTGGATCCACTTGAA 270
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98256 CTGAAGACCACGATGTCTGTTTCTAGAAATGAGTTGTTGCAGAAGTTGGATCCACTTGAA 98197

Qy    271 CAAGCAAAAGTGGATTTGGTTTCTGCATACACACATTAAATTCAAGTTTGGGTTTATTTG 330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98196 CAAGCAAAAGTGGATTTGGTTTCTGCATACACACATTAAATTCAAGTTTGGGTTTATTTG 98137

Qy    331 GCAACCCAAGGAGTTAATCCTAGGGAACACATCCAGTAAAACAGGAATTGGAAAGAATCAGA 390
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98136 GCAACCCAAGGAGTTAATCCTAGGGAACACATCCAGTAAAACAGGAATTGGAAAGAATCAGA 98077

Qy    391 GTATATATGAACAGAGAGCTCAGGGAAATAACAGACAGACAAAGAAAAGGCTGGCAAGCTGGACAGA 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98076 GTATATATGAACAGAGAGCTCAGGGAAATAACAGACAAGAAAAGGCTGGCAAGCTGGACAGA 98017

Qy    451 GGTGCAGCTTCAAGATTTGTAAAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAATGCATCA 510
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98016 GGTGCAGCTTCAAGATTTGTAAAAAAAATGCCCTCTGGGAACCAAAAACCGAAAAATGCATCA 97957

Qy    511 AAAGTTGCCAATAAAGGAAAAGTAAACTTAACTTTTTTGGTTTTGGATGTACACATATTC 570
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97956 AAAGTTGCCAATAAAGGAAAAGTAAACTTAACTTTTTTGGTTTTGGATGTACACATATTC 97897

Qy    571 AAAAAAGTAC------------------------------------------- 579
          |||||||||
Db  97896 AAAAAAGTACATTCCCCGCGCCCCCCCCACAGAATAATTCTGTGGCAGGACAAGG 97837

Qy    580 -------ATTAATAAGTAATCACAG---TAATATGTAAAGCTAAATAC-TT 619
          |||||||||||||||||||||||        |||||||||||||||  ||
Db  97836 TTTAAATGTGTTTCTTATTAATATGTAAAATTCACAGTAAAATCACAGTAAAAATACTTT 97777

Qy    620 CCTCTCCAAAGATCATTATATCTTTATTGACTGATTGGACTGAGGATTTTAACATTGTGATATAT 679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97776 CCTCTCCAAAGATCATTATCTTTATTGACTGATTGGACTGAGGATTTTAACATTGTGATATAT 97717

Qy    680 TATATATTTTATAATTTTAATTTACCCATCTC---TTGATGAGACTCCTTATTCTTTATTATAAGGTCAGT 737
          |||||||||||||||||||||||||||||        |||||||||||||||||||||||
Db  97716 TATATATTTTATAATTTTAATTTACCATCTCTTTTTGATGAGACTCCTTATTCTTTATTATAAGGTCAGT 97657

Qy    738 CTTTGCAAGTACCATTTTTATTAAGCAGCTGTGAAATTTAAGTGAAATGTTCAGAATTAT 797
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97656 CTTTGCAAGTACCCATTTTTATAGCAGCTGTGAAATTTAAGTGAAATGTTCTTTGTAAACAT 97597

Qy    798 TTGTACTATTTTTAAATGAAATAAATGACCTTATGTAGGCTGTGAAATTAT 857
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97596 TTGTACTATTTTTAAATGAAATAATGACCTTATGTAGGCTGTGAAATTAT 97537

Qy    858 AGGTACATCTGTTTTCACTATATGATATTAAGAAAAGCGTG-AAGACTTAAAATGTTCATT 916
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  97536 AGGCACATCTGTTTTCACTATATGATATTAAGAAAAGCATGAAAGACTTAAAATGTTCATT 97477

Qy    917 TTTTTCTGTATAGAGATACTTTATCATGATTTTTCATTGATTACTGCTTTGTGTGATGAT 976
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97476 TTTTTCTGTATAGAGATACTTTATCATGATTTTTCATTGATTTACTGCTTTGTGTGAT 97417

Qy    977 ATTCAAAGTGTGAAAACTAAAAGTTTATATGGTTGTGTACTTTTAAATTCTTGGGCCATGTTGCCTCTA 1036
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97416 ATTCAAAGTGTGAAAACTAAAAGTTTATGTGGTTGTGTACTTTTAAATTCTTGGGCCATGTTGCCTCTA 97357

Qy   1037 TGTCCCATTTAAAATAAAATACAATTCTCATTAACTTTAGATGGGAAATAAGGTTGTAGT 1096
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97356 TGTCCCATTTAAAATAAAATACAATTCTCATTAACTTTAGATGGGAAATAAGGTTGTAGT 97297

Qy   1097 TGATGGATGAAATTTGGCATGATGAGACTGTACTCTCAATAAAGGCTGAAAATGTTGTATAA 1156
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97296 TGATGGATGAAATTTGGCATGATGAGACTGTACTCTCAATAAAGGCTGAAAATGTTGTATAA 97237
```

RESULT 13

AC079112/c

LOCUS   AC079112   155888 bp   DNA   linear   PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-67K11 from 2, complete sequence.
ACCESSION   AC079112
VERSION   AC079112.4 GI:14091934
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 155888)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2 (bases 1 to 155888)
  AUTHORS   Goyea,E., Mishra,S. and Kozlowicz,A.
  TITLE   The sequence of Homo sapiens BAC clone RP11-67K11
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 155888)
  AUTHORS   Waterston,R.H.
  JOURNAL   Direct Submission
            Submitted (18-AUG-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 155888)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5 (bases 1 to 155888)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On May 16, 2001 this sequence version replaced gi:13752170.
COMMENT   -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_NH0067K11

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis,
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome

```
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-30P5; the clone sequenced
to the right is RP11-474G23. Actual start of this clone is at base
position 1 of RP11-67K11; actual end is at base position 155888 of
RP11-67K11.

FEATURES             Location/Qualifiers
source               1..155888
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-67K11"
                     /clone_lib="RPCI-11"

repeat_region        328..526
                     /rpt_family="MER1_type"
repeat_region        538..645
                     /rpt_family="MIR"
repeat_region        1745..2080
                     /rpt_family="ERV1"
repeat_region        2082..2215
                     /rpt_family="ERV1"
repeat_region        2219..2389
                     /rpt_family="Alu"
repeat_region        2390..2821
                     /rpt_family="ERV1"
repeat_region        3746..3978
                     /rpt_family="ERV1"
repeat_region        4181..4281
                     /rpt_family="MIR"
repeat_region        4536..4849
                     /rpt_family="MIR"
repeat_region        5603..5990
                     /rpt_family="Alu"
repeat_region        5993..7584
                     /rpt_family="MaLR"
repeat_region        7585..7708
                     /rpt_family="MaLR"
repeat_region        7709..7824
                     /rpt_family="Alu"
repeat_region        7825..7905
                     /rpt_family="MaLR"
repeat_region        7907..8614
                     /rpt_family="MaLR"
repeat_region        8615..10087
                     /rpt_family="L1"
repeat_region        10088..10108
                     /rpt_family="L1"
repeat_region        10172..10432
                     /rpt_family="AT_rich"
repeat_region        10433..10741
                     /rpt_family="Alu"
repeat_region        11491..11646
                     /rpt_family="MaLR"
repeat_region        12515..12833
                     /rpt_family="MIR"
repeat_region        12910..13181
                     /rpt_family="Alu"
repeat_region        13182..13483
                     /rpt_family="Alu"
repeat_region        13749..14156
                     /rpt_family="Alu"
repeat_region        14158..14654
                     /rpt_family="L2"
repeat_region        14763..14928
                     /rpt_family="MaLR"
repeat_region        14990..15229
                     /rpt_family="MaLR"

repeat_region        15940..17923
                     /rpt_family="L1"
repeat_region        18057..18200
                     /rpt_family="L1"
repeat_region        18806..18917
                     /rpt_family="L1"
repeat_region        19995..19331
                     /rpt_family="MIR"
repeat_region        20065..20084
                     /rpt_family="AT_rich"
repeat_region        20091..20135
                     /rpt_family="(TC)n"
repeat_region        20136..20179
                     /rpt_family="(TG)n"
repeat_region        21523..21686
                     /rpt_family="CT-rich"
repeat_region        21829..22040
                     /rpt_family="MER1_type"
repeat_region        22370..22664
                     /rpt_family="MIR"
repeat_region        22757..22858
                     /rpt_family="L1"
repeat_region        23117..23428
                     /rpt_family="Alu"
repeat_region        23565..24609
                     /rpt_family="L1"
repeat_region        25098..25143
                     /rpt_family="L1"
repeat_region        25479..25657
                     /rpt_family="AT_rich"
repeat_region        25658..25929
                     /rpt_family="L2"
repeat_region        25930..26003
                     /rpt_family="Alu"
repeat_region        26003..26168
                     /rpt_family="L2"
repeat_region        26390..26739
                     /rpt_family="MIR"
repeat_region        26789..27175
                     /rpt_family="MER2_type"
repeat_region        27258..27303
                     /rpt_family="MaLR"
repeat_region        27536..27604
                     /rpt_family="(CA)n"
repeat_region        27589..27627
                     /rpt_family="MIR"
repeat_region        28276..28369
                     /rpt_family="L2"
repeat_region        28432..28836
                     /rpt_family="MIR"
repeat_region        28950..30197
                     /rpt_family="L2"
repeat_region        30283..30327
                     /rpt_family="MIR"
repeat_region        30417..30565
                     /rpt_family="(CA)n"
repeat_region        31120..31275
                     /rpt_family="ERVL"
```

```
Query Match    56.7%;  Score 655.8;  DB 9;  Length 155888;
Best Local Similarity   90.7%;  Pred. No. 6.4e-108;
Matches 772; Conservative 0; Mismatches 7; Indels 72; Gaps 4;

Qy     378 GGAAAGAAATCAGAGTATATATGAACAGAGTCAAGAGAAATAACAGAACAGAAAAAGGCTGG 437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116453 GGAAAGAAATCAGAGTATATATGAACAGAGTCAAGAGAAATAACAGAACAGAAAAAGGCTGG 116394

Qy     438 CAAGCTGGACAGAGGTGCCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGAACCAAAAATC 497
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116393 CAAGCTGGACAGAGGTGCCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGAACCAAAAATC 116334

Qy     498 GAAAAATGCATCAAAAAGTTGCCAATAAAAGGAAAAAAGTAAAAGTTAACTTTTTTGGTTTTGA 557
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116333 GAAAAATGCATCAAAAAGTTGCCAATAAAAGGAAAAAAGTAAAAGTTAACTTTTTTGGTTTTGA 116274
```

```
Qy    558 TGTACACATATTCAAAAAGTAC-----------------------------------------------------579
          ||||||||||||||||||||||
Db 116273 TGTACACATATTCAAAAAGTACAACATTCCCCCCCCCCCCCCCGCAAAAATAATAATTCTG 116214

Qy    580 -------------------------------ATTAAATATAGTAATCACA---GTAAATATNGTAA 607
                                         |||||||||||  |||||
Db 116213 TGGCACGGCAAGGTTAAATGTNGTTTCTTATTAAATATNGTAAAATTCACAGTAAATATNGTAA 116154

Qy    608 AGCTAAATAC-TTCCCTCCAAAGATCATTATCTTTATTTTATTGCATTGGACACTGCAGGATTTTAA 666
          ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 116153 AGCTAAATACTTCCCTCCAAAGATCATTATCTTTATTTTATTGCATTGGACACTGCAGGATTTTAA 116094

Qy    667 CATTGTGATATATATTATTATATTTTTATTTTTAATTTTACCAACTCTTATTTCTTTA 726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 116093 CATTGTGATATATATTATTATATTTTTATTTTTAATTTTACCAACTCTTATTTCTTTA 116034

Qy    727 TATAGGTCAGTCTTCTTGCCAAGTACCATTTATAATTTAAAGCAGCTGTGAAATTTATTCTTTA 786
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 116033 TATAGGTCAGTCTTCTTGCCAAGTACCATTTATAAGAGCTGTGAAATTTAAGTGAAATGTTTC 115974

Qy    787 TTTGTAAACATTTGTACTATTTTAAATGAATAAGACCTTATAGATATGCTATCCTGTAG 846
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 115973 TTTGTAAACATTTGTACTATTTTAAATGAATAAGACCTTATAGATATGCTATCCTGTAG 115914

Qy    847 GCTCGAAATTATAGGTACATCTGTTTTCACTATATGATAATTAAGAAAAGGGTG-AATGACTT 905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db 115913 GCTCGAAATTATAGGTACATCTGTTTTCACTATATGATAATTAAGAAAAGGGTGAAATGACTT 115854

Qy    906 AAATGTTCATTTTTTTCCGTATAGATACTTTACTCATGTTTCATGAAGAATTACT 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 115853 AAATGTTCATTTTTTTCGTATAGATACTTTACTCATGTTTCATGAAGAATTACT 115794

Qy    966 GCTTTGTTGATATTCAAAGTGTGAAACTAAAAGTTTTATGGTTGTGTACTTTAATTCTTGGCA 1025
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 115793 GCTTTGTTGATATTCAAAGTGTGAAACTAAAAGTTTTATGGTTGTGTACTTTAATTCTTGGCA 115734

Qy   1026 TGTTGCCCTATCTGTCCCATTTAAAAATAAAATAACATTCTCATTAAACTTTAGATGGGAAATA 1085
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 115733 TGTTGCCCTATGTCCCATTTAAAAATAAAATAACATTCTCATTAAACTTTAGATGGGAAATA 115674

Qy   1086 AGGTTGTATGTTGATGGATGAATTTTGGCATGATGACTGTACTCTCAAATAAAGGCTGAAA 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 115673 AGGTTGTATGTTGATGGATGAATTTTGGCATGATGACTGTACTCTCAATATAAGGCTGAAA 115614

Qy   1146 ATGTTGTAAA 1156
          ||||||||||
Db 115613 ATGTTGTATAA 115603
```

RESULT 14
AL356095/c

```
LOCUS       AL356095      172079 bp    DNA     linear   PRI 24-JAN-2001
DEFINITION  Human DNA sequence from clone RP11-369J21 on chromosome 10,
            complete sequence.
ACCESSION   AL356095
VERSION     AL356095.11  GI:12331029
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172079)
  AUTHORS   Brown,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 22, 2001 this sequence version replaced gi:11863400.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations,
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-369J21 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-479017 is at 65134 in this sequence.
            right end of clone RP11-369J21 The true
FEATURES             Location/Qualifiers
     source          1..172079
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-369J21"
                     /clone_lib="RPCI-11.2"
     misc_feature    7..431
                     /note="match: GSS: Em:AQ191918"
     misc_feature    25..413
                     /note="match: GSS: Em:AQ815029"
     misc_feature    61..595
                     /note="match: GSS: Em:AQ531010"
     repeat_region   647..944
                     /note="L1PA12 repeat: matches -1422..-1118 of consensus"
     repeat_region   1339..1734
                     /note="L2 repeat: matches 1652..2042 of consensus"
     repeat_region   2037..2348
                     /note="AluSx repeat: matches 1..312 of consensus"
     repeat_region   2389..2697
                     /note="AluSx repeat: matches 1..312 of consensus"
     repeat_region   2939..3074
                     /note="AluSx repeat: matches 2..310 of consensus"
     repeat_region   3197..3411
                     /note="FLAM_C repeat: matches 12..141 of consensus"
     repeat_region   3518..3821
                     /note="MER58A repeat: matches 1..224 of consensus"
     repeat_region   4024..4580
                     /note="AluSx repeat: matches 1..300 of consensus"
     repeat_region   4626..4766
                     /note="L2 repeat: matches 2210..2750 of consensus"
     repeat_region   4768..5076
                     /note="L2 repeat: matches 2240..2383 of consensus"
     repeat_region   5475..5997
                     /note="AluSx repeat: matches 1..312 of consensus"
     repeat_region   5994..6157
                     /note="L1MCb repeat: matches 102..831 of consensus"
     repeat_region   6158..6294
                     /note="L1M4 repeat: matches 4140..4298 of consensus"
     repeat_region   6295..6750
                     /note="FLAM_C repeat: matches 1..133 of consensus"
     repeat_region   6751..7063
                     /note="L1M4 repeat: matches 4298..4741 of consensus"
     repeat_region   7064..7628
                     /note="AluY repeat: matches 1..311 of consensus"
     repeat_region   7629..7915
                     /note="L1M4 repeat: matches 4741..5310 of consensus"
     repeat_region   7916..8047
                     /note="L1M4 repeat: matches 1..284 of consensus"
     repeat_region   8054..8180
                     /note="AluSx repeat: matches 5310..5437 of consensus"
```

```
repeat_region    8182. .8411    /note="FLAM_C repeat: matches 2. .140 of consensus"
repeat_region    8412. .8717    /note="L1PA7 repeat: matches 5432. .5661 of consensus"
repeat_region    8718. .8862    /note="L1PA7 repeat: matches 5661. .5805 of consensus"
repeat_region    8863. .9012    /note="L1PA7 repeat: matches 5992. .6141 of consensus"
repeat_region    9140. .9288    /note="L1PA7 repeat: matches 151. .6141 of consensus"
repeat_region    9292. .9516    /note="AluJo repeat: matches 1. .151 of consensus"
repeat_region    10088. .10389  /note="L1MB6 repeat: matches 5971. .6182 of consensus"
repeat_region    10390. .10697  /note="AluSq repeat: matches 1. .302 of consensus"
repeat_region    11104. .11317  /note="Aluy repeat: matches 1. .308 of consensus"
repeat_region    11749. .12047  /note="MER46A repeat: matches 22. .235 of consensus"
repeat_region    12418. .12472  /note="AluSg repeat: matches 1. .299 of consensus"
repeat_region    12473. .12772  /note="L1MB2 repeat: matches 3968. .4023 of consensus"
repeat_region    12773. .13104  /note="AluSg repeat: matches 1. .299 of consensus"
repeat_region    13105. .13308  /note="L1MB2 repeat: matches 4023. .4362 of consensus"
repeat_region    13409. .13722  /note="AluSg repeat: matches 1. .303 of consensus"
repeat_region    13410. .14039  /note="AluY repeat: matches 1. .310 of consensus"
repeat_region    14040. .14344  /note="L1MB2 repeat: matches 4362. .4638 of consensus"
repeat_region    14345. .14683  /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region    14745. .15005  /note="L1MB2 repeat: matches 4638. .4985 of consensus"
misc_feature     complement(14996. .15536)  /note="match: GSS: Em:AQ356820"
misc_feature     complement(15014. .15524)  /note="match: GSS: Em:AQ192026"
repeat_region    15137. .15273  /note="AluSp repeat: matches 1. .291 of consensus"
repeat_region    15310. .15539  /note="AluJo repeat: matches 1. .134 of consensus"
repeat_region    15540. .15828  /note="L1MB2 repeat: matches 5042. .5257 of consensus"
repeat_region    15829. .16547  /note="AluJb repeat: matches 3. .291 of consensus"
repeat_region    16558. .16666  /note="L1MB2 repeat: matches 5257. .6013 of consensus"
repeat_region    16675. .16979  /note="MLT1B repeat: matches 2. .121 of consensus"
repeat_region    16994. .17094  /note="AluSx repeat: matches 1. .306 of consensus"
repeat_region    complement(17009. .17406)  /note="L1MB2 repeat: matches 6069. .6168 of consensus"
misc_feature     17971. .18031  /note="match: GSS: Em:AQ076012"
repeat_region    18034. .18072  /note="Alu repeat: matches 2. .62 of consensus"
repeat_region    18073. .18375  /note="MER4D repeat: matches 564. .601 of consensus"
repeat_region    18376. .18422  /note="AluJb repeat: matches 2. .301 of consensus"
repeat_region    18490. .18634  /note="MER4D repeat: matches 601. .651 of consensus"
repeat_region    18635. .18895  /note="LTR29 repeat: matches 53. .198 of consensus"
repeat_region    18947. .19130  /note="MER41B repeat: matches 1. .271 of consensus"
repeat_region    19130. .19930  /note="AluSg/x repeat: matches 109. .292 of consensus"

repeat_region    19128. .19305  /note="AluSq repeat: matches 140. .313 of consensus"
repeat_region    19306. .19377  /note="36 copies 2 mer aa 75% conserved"
repeat_region    19723. .19808  /note="LTR8 repeat: matches 1. .73 of consensus"
repeat_region    19826. .19975  /note="LTR8 repeat: matches 137. .286 of consensus"
repeat_region    19976. .20597  /note="AluJb repeat: matches 60. .689 of consensus"
misc_feature     complement(20789. .21658)  /note="match: GSS: Em:AQ891390"
repeat_region    20900. .21045  /note="AluJo repeat: matches 1. .127 of consensus"
repeat_region    21046. .21335  /note="AluSx repeat: matches 6. .297 of consensus"
repeat_region    21336. .21501  /note="AluJo repeat: matches 127. .299 of consensus"
repeat_region    22184. .22478  /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region    22687. .22768  /note="HERVL repeat: matches 2. .84 of consensus"
repeat_region    23426. .23535  /note="AluS repeat: matches 1. .98 of consensus"
repeat_region    23536. .23826  /note="HERVL repeat: matches 1. .294 of consensus"
repeat_region    23827. .23869  /note="AluSg repeat: matches 1. .294 of consensus"
repeat_region    23932. .24283  /note="HERVL repeat: matches 98. .143 of consensus"
repeat_region    24284. .24540  /note="AluSx repeat: matches 4267. .4622 of consensus"
repeat_region    24541. .24787  /note="AluSx repeat: matches 1. .294 of consensus"
```

```
Query Match    44.9%;  Score 519.2;  DB 9;  Length 172079;
Best Local Similarity   84.0%;
Matches  675;  Conservative   0;  Mismatches  58;  Indels  71;  Gaps  5;

Qy      91 GGAGGGTGAGGAGCAGAGCTGGCCATAAATGGCCAGGTGAAGAAATTAATGAAGAACTATCCA 150
           |||  | ||||| | || ||||| ||| |||| || |||||||||||||||| ||| |||||
Db   36364 GGAGTCGGAAGGCCGTGAGTCAGCCACATGAGTCGCAGCTGAAGAAATTAATGAGGAGTATCCA 36305

Qy     151 GTAGAAATTCACGAGTATTTGTCAGCGTTTCAGAAGAATCCATTGGTGCTGTGGATGAGATG 210
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36304 GTAGAAATTCACGATTATTTTGTCAGCATTTGCGAATTCCATTGATGCTGTGGATGAGATG 36245

Qy     211 CTGAAGACCATGACTGTCTGTTTCTAGAAGATGAGTTGTTGCAGAAGTTGGATCCACTTGAA 270
           ||||| ||| | || | |||  ||| ||| ||| ||| |||||||| ||||||  ||||||
Db   36244 CTGAAGAGAACATGAGTGTCGTTTCTAGAAATGAGTTGTTGCAGAAGTTGGACCCACTTGAA 36185

Qy     271 CAAGCAAAAGCTGGGATTTTGGTTTTCTGCATACACATTAAATTCAAATCTGTTTCGGGTTTATTTG 330
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36184 CAAGCAAAAAGTGGGATTTTGGTTTTCTGCATACACATTAAATTCAAATCGTTTTCGGGTTTATTTG 36125

Qy     331 GCAACCCAAGGAGTTAATCCTAAGGAACATCCAGTAAAACAGGAGTTGGAAGAATCAGA 390
           |||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||
Db   36124 GCAACTCAAGGAGTGAATCCTAAGGAACATCCAGTAAAAGCAGGAGTTGGAAGAATCAGA 36065

Qy     391 GTATATATGAACAGAGTCAAGGAAATAACAGACAAGAAAAAGGCTGGCCAAGCTGGACAGA 450
           ||||||||||||||||||||||||||||||||||||| | |||||||||||||||||||
Db   36064 GTATATATGAACAGAGTCAAGGAAAATAACAGACAAGAAAAAGGCTGGCCAAGCTGGACAGA 36005

Qy     451 GGTGCAGCTTCAAGATTTTGTAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAATCGATCA 510
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36004 GGTGCAGCTTCAAGATTTTGTAAGAAAATGCCCTCTGGGAACCAAAACGGAAAAATACATCC 35945

Qy     511 AAAGTTGCCAATAAAGGAAAAAGTAAAAGTTAACTTTTTGGTTTTGAGTGTACACATATTC 570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35944 AAAGTTGCCCATAAAGGAAAAAGTAAAAGTAACTTTTTGGATTTGATGTACACATATTC 35885

Qy     571 AAAAAGTACA 589
           |||||||| |
Db   35884 AAAAAGTACATCTCTTTTTTTCCCAAAAAGTAATTCTGTGACAGAGACCAAGGTTTAAATGT 35825
```

RESULT 15

LOCUS   BC005436   2000 bp   mRNA   linear   ROD 29-JUN-2004
DEFINITION Mus musculus nuclear DNA binding protein, mRNA (cDNA clone MGC:5983
            IMAGE:3591682), complete cds.
ACCESSION   BC005436
VERSION   BC005436.1 GI:13529391
KEYWORDS   MGC.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 2000)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2 (bases 1 to 2000)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 11 Row: d Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 10181123.

FEATURES   Location/Qualifiers
     source   1..2000
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="MGC:5983 IMAGE:3591682"
                     /tissue_type="Mammary tumor metastatized to lung. Tumor
                     arose spontaneously from a senescent normal mammary
                     (clonal) outgrowth infected with the virus MMTV."
                     /clone_lib="NCI_CGAP_Lu29"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene   1..2000
                     /gene="Cid"
                     /note="synonym: SUN-CoR"
                     /db_xref="LocusID:57316"
                     /db_xref="MGI:1927354"
     CDS   47..472
                     /gene="Cid"
                     /codon_start=1
                     /product="nuclear DNA-binding protein"
                     /protein_id="AAH05436.1"
                     /db_xref="GI:13529392"
                     /db_xref="LocusID:57316"
                     /db_xref="MGI:1927354"
                     /translation="MAGEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITD
                     KKKAAKLDRGAASRFVKNALWEPKAKSTPKVANKGKSKH"

ORIGIN
Query Match   40.1%; Score 464; DB 10; Length 2000;
Best Local Similarity   71.5%;
Matches 719; Conservative   0; Mismatches 255; Indels   32; Gaps   7;

Qy   106 GAGCTGGCCATAATGGCAGGTGAAGAAAATTAATGAAGACTATCCAGTAGAAATTCACGAG 165
               |||||| ||| || |||||||| ||||| ||||| || | |||| |||||||||||||||
Db   35 GAGTTGGTCACAATGGCAGGTGAAGAAATGAAGAGTTAATCCCGTAGAAAATTCACGAG 94

Qy   166 TATTTGTCAGCGTTTGAGAATTCCATTGGTGTCGTGTAGATGAGAGATGCTGAAGAGACCATATG 225
               |||| |||| ||||||||||| || | || ||| ||| || ||||||||||| | ||| ||
Db   95 TCTTTAACAGCCCCTGGAGAGCTCCCCTGGGTGCGTGGAGATGACAATGCGAAGACCATATG 154

Qy   226 TCTCGTTTCTAGAAATGAGTTGTTGGCAGAAGTTGGATCCACTTGAACAAGACAAAAGTGGAT 285
               ||||||||| |||||||| ||| ||||| ||||||||| || |||||||| |||||||||||
Db   155 GCTCGTTTCTTAGAAATGAGTTGTTGCAGAAGTTGGACCCATTGAACAAGACAAAAGGTGGAT 214

Qy   286 TTGGTTTCTGCATACACATTAAAATTCAATGTTTGGGTTTATTGGCAACCCAAGGAGTT 345
               ||| |||||| |||| | |||| | | |||| ||| |||||||||||||||| || |||
Db   215 TTAGTTTCTGCATACACCTTAAATTCAATGTTTTGGGTTTATTGGCAACTCAAGGAGTT 274

Qy   346 AATCCTAGGACACATCCAGTAAAACAGGAATTGGAAGAAATCAGAGTATATATGAACAGA 405
               ||| |||| | || |||||||| | |||||||||||| || ||||| || | |||||||
Db   275 AATCCCAAAGAGCATCCAGTGAAGCAGGAACTGGAAAGAATCAGAGTCTACACATGAACAGA 334

Qy   406 GTCAAGGAGAAATAACAGACAAGAAAAAAGGCTGGCAACGTGGACAGAGAGGTGCAGCTTCAAGA 465
               || ||||| || ||| |||||||||||| |||| ||| || ||| ||| |||| ||| |||||
Db   335 GTTAAAGAAAATAACAGACAAGAAGAAGAGGCTGCCAAGCTGGACAGAGGAGCTGCTTCGAGA 394

Qy   466 TTTGTAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAAGTGCATCAAAAGTTGCCAATAAA 525
               ||| |||| || ||| || || ||||||| | | |||||| ||| | | |||||||||| |
Db   395 TTTGTCAAGAACGCACTCTCGGAACCCAAAGAACACACAAAAAAGCACACCAAAAAGTACATTAA 454

Qy   526 GGAAAAAGTAAAAGTTAACTTTTGGTTTGATGTACACATATTCAAAAAGTACATTAAT 585
               |||||||||| ||||| | || ||| |||| || | ||| || ||||| | |||||
Db   526

Qy   590 AATCACAGT----------------AATATGTAAAGCTAAATAC-TTCCTC 623
               |||||| |||||||| | |||||| ||||||
Db   35824 ATTCCTTGTTGTGAATTAATATGTAAAATTACAGTAAATTTGTAAAGCTCAATACTTTCCTC 35765

Qy   624 TCCAAAGATCATATCTTTATGATTGATTGACAC--TGAGGAGTTTTAACATTGTGATATATTA 681
               |||||| ||| | || |||||| || ||| |||| |||||| ||||| || || |||
Db   35764 TCCAAAGGTTGTTATCTTTTATTGATTGCACTATGAGGGGATTTTAACATTGTGATATATTT 35705

Qy   682 TATATTTTATATTAATTTTACCATCTC--TTGATGAGACTTCTTATTTCTTTTATATAAGGTCAGTCT 739
               ||| |||| |||||| | |||| ||| |||||||| | | || ||||| | || |||||||||||
Db   35704 TATATTTTACAATTTACCATCTCTTTATTTTTTTGAAGAGACTTCTTATTTCTTTTATATAAGGTCAGTCT 35645

Qy   740 TGCAAGTACCATTTTATAAGCAGCTGTGTGAAAATTTTAAGTTCAATGTTCTTTGTAAACATTT 799
               |||| |||| || |||||||||| |||| || || ||||||| || ||||||| | || ||||
Db   35644 TGCAGGTAGCATTTTCTTAGGAACTGTGTGAAAATTTTAAGTGAAAAGTTCTTTGAAAATATTT 35585

Qy   800 GTACTATTTTAAATGAATAATGAC 823
               |||||||||| || ||| ||||||
Db   35584 GTACTATTTTCAATGAAAATAACAC 35561

```
Db    455  GGGAAAAGCAAAACACTAATCTTTTGGTTTTGATGTGACATGTTTCAAAAAGTACA-TCCT  513
            |||  |||        ||||||||   |||||| || |||||||   |||||||||||| |

Qy    586  ATGTAATCACAGTAATATGTAAAGCTAAATACTTCCTCCTCCAAAGATCATTATCTTTATT  645
            ||||| |||        |||||| ||||  |||

Db    514  TTTTAATCAGTTTACAATGTAGTTAATGTGACCAATGTGGTGTTTAAATGGATTCCTTTTGG  573

Qy    646  GAT-----TAGCACTGAGGATTTTAACATTGTGATATATATATATATTTATATTTACCAT  700
            |  |||      ||| || ||||||| || ||||||||||||||||||||||||||||

Db    574  AATTCATGTATAAAATTACACATTACATTTGTGATACTGAAATCTTTTTTTTGCTGAGAAA  633

Qy    701  CTCTTGAGTGAGGACTCTTATTTCTTTATATAGGTCAGTCTTGCAAGTACCATTTTATAAGC  760
            |||     ||         |||  |||||||||

Db    634  GATTAAGTTGTGTCTTTGTTGATTTTCATATAAAGCATCATGATGTGTTTAATATTGTTAAGA  693

Qy    761  AGCTGTGAAA-------TTTAAGTGAAAAGTTTCTTTGTAAACATTTGTACTATTTTTAAAT  813
            |||||||||              ||||||| |||||| |||||||||||  ||||||

Db    694  TATTCTATAAGCAGTTGTGAAATCCAAATGTTCTCTGTAAACATTTGTAGTGTTTGAAAAT  753

Qy    814  GAATAAATGACCTTATGAAGTATGCTATCTCGTGTAGGCTGAAATTATAGGTACATCTGTTTTC  873
            |||         ||||||| ||||||||  ||||||           |||  |||||||

Db    754  GAACAATGATATTATGAAGTGTGCTATCTCGTGTAGACTGGAGGTGTGAAGGACATTTGTTTTC  813

Qy    874  ACTATATGATATTTAAGAAAG---CGTGAATGACTTAAATGTTCATTTTTTCTGTATAGA  930
            | |           |||||      |||||||| ||||| |||||| |||||||||||

Db    814  AGTAATGATGAGAAATACAGTGACTTAAATACCCACTCTGTTTTCTGTTCAGTTAGTTCA  873

Qy    931  TACTTTATCATGTTTTCATGATTTTAGGAATTACTGCTTTTGTTGATATTCAAAGTGTGAA  990
            |||||  ||||| |||||| ||| |||||||||| || ||||||||||||   ||||||

Db    874  ACATGTTTCTGTGATTTTTTTTTTGGGAGTAATTGCTGTCTTGATAATTCAAAGTACAAA  933

Qy    991  ACTAAAAGTTTATGGTTGTACTTTAATTCTTGGCATTGTGCCTCTATGTCCCATTTAAAA  1050
            |  |||     |||  |||||||||||||||  |||||| ||||| ||||||||||||

Db    934  ATTGAAAACCTTAAGGCTGTTACTTTAAATTCTT-------------CATGTTCCATTTAAAA  980

Qy    1051 TAAAATACATTCTCATTAACTTTAGATGGGAAATAAGGTTGTATGT  1096
            |||||          ||||||||  || |||||||||||| |||||

Db    981  TAAAAT--GTTCTCATTAACTCT-GATGGAAAACACGGTTGTATGT  1023
```

This Page Blank (uspto)

```
OM protein - protein search, using sw model

Run on:          January  4, 2005, 08:13:30        GenCore version 5.1.6
                                                   Copyright (c) 1993 - 2005 Compugen Ltd.

Title:           US-09-701-618A-2
Perfect score:   715
Sequence:        1 MAGEEINEDYPVEIHEYLSA.........LWEPKSKNASKVANKGKSKS 141

Scoring table:   BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:        2002273 seqs, 358729299 residues
                                                   766.376 Million cell updates/sec
                                                   Search time 66 Seconds
                                                   (without alignments)

Total number of hits satisfying chosen parameters:    2002273

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  A_Geneseq_23Sep04:*
      1: geneseqp19808:*
      2: geneseqp19908:*
      3: geneseqp20008:*
      4: geneseqp20018:*
      5: geneseqp20028:*
      6: geneseqp2003aa:*
      7: geneseqp2003bs:*
      8: geneseqp20048:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	715	100.0	141	3	AAY51024	Human C1D
2	715	100.0	141	7	ADE59764	Human Pro
3	671	93.8	141	7	AAO15405	Mammalian
4	667	93.3	140	5	ADI28083	Human nuc
5	654	91.5	141	7	ADE59762	Rat Prote
6	646	90.3	141	5	AAY51025	Murine C1
7	644	90.1	140	5	AAO15406	Mammalian
8	519	72.6	102	3	AAG01231	Human sec
9	266	37.2	54	8	ABO54730	Human gen
10	150.5	21.0	217	3	AAG49051	Arabidops
11	150.5	21.0	217	3	AAG24127	Arabidops
12	150.5	21.0	253	3	AAG49050	Arabidops
13	150.5	21.0	256	3	AAG24126	Arabidops
14	141.5	19.8	177	3	AAG24128	Arabidops
15	141.5	19.8	177	3	AAG49052	Arabidops
16	126.5	17.7	159	4	ABB64082	Drosophil
17	119.5	16.7	232	8	ADP99133	C. albica
18	110	15.4	184	7	ABR53602	Protein s
19	110	15.4	184	7	ADK64282	Disease t
20	110	15.4	151	7	ADH88594	Enterococ
21	85	11.9	1294	4	ABB63502	Drosophil
22	83	11.6	1249	6	ABU44466	Protein e
23	82	11.5	345	3	AAG23880	Arabidops
24	81.5	11.4	295	3	ADH85852	Enterococ
25	81.5	11.4	307	7	ADC97413	E. faeciu
26	81.5	11.4	904	2	AAY16123	Mycoplasm
27	80.5	11.3	675	6	ABU43960	Protein e
28	79	11.0	479	3	AAY92242	Human can
29	79	11.0	479	5	ABO55690	Human nuc
30	79	11.0	479	5	ABB83498	Large hum
31	79	11.0	479	5	ABP64721	Human pro
32	79	11.0	479	4	ABP69180	Human hea
33	79	11.0	948	4	ABB61027	Drosophil
34	78.5	11.0	152	8	ADN46815	Thermococ
35	78.5	11.0	703	6	ABU36173	Protein e
36	78	10.9	508	6	AAY92660	Mutant mu
37	78	10.9	923	4	AAY92662	Mutant mu
38	78	10.9	1094	2	ABP73717	Drosophil
39	77.5	10.8	2285	2	AAW98149	Bacillus
40	77.5	10.8	209	6	ABM67922	Photorhab
41	77	10.8	465	4	AAU34291	Candida a
42	77	10.8	465	2	AAW22467	Staphyloc
43	77	10.8	466	2	AAW58529	Staphyloc
44	77	10.8	466	3	AAY91053	Staphyloc

ALIGNMENTS

RESULT 1
AAY51024

```
ID   AAY51024 standard; protein; 141 AA.
XX
AC   AAY51024;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Human C1D protein.
XX
KW   C1D; human; apoptosis; tumour; gene therapy; treatment.
XX
OS   Homo sapiens.
XX
XX
PN   DE19824811-A1.
XX
PD   09-DEC-1999.
XX
PF   03-JUN-1998;  98DE-01024811.
XX
PR   03-JUN-1998;  98DE-01024811.
XX
PA   (DEKR-)  DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Rothbarth K,  Stammer H,  Werner D;
XX
DR   WPI; 2000-063506/06.
DR   N-PSDB; AAZ43927.
XX
PT   Inducing apoptosis by overexpressing the C1D gene, particularly for
PT   treating tumors.
XX
PS   Claim 3; Fig 1; 10pp; German.
XX
CC   This invention describes a novel method for inducing apoptosis which
CC   comprises overexpressing the C1D gene (I). The method is particularly
CC   used to treat tumors and can also be used in gene therapy. The method has
CC   no side effects on normal cells (contrast known methods of inducing
CC   apoptosis such as cytotoxins and radiation), and may be effective on
CC   cells resistant to conventional treatments. Overexpression of (I) is
CC   sufficient itself to induce apoptosis but the effect may be increased
CC   when used in combination with other anti-tumor methods. When cells
CC   transfected with (I) undergo apoptosis, they release factors that kill
CC   neighboring, non-transfected cells (bystander effect). This sequence
CC   represents the human C1D protein described in the method of the invention
XX
SQ   Sequence 141 AA;
```

```
Query Match          100.0%;  Score 715;  DB 3;  Length 141;
Best Local Similarity 100.0%;  Pred. No. 1.9e-66;
Matches  141;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITDKKKAGKLDRGAASRFVKN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITDKKKAGKLDRGAASRFVKN 120

Qy  121 ALWEPKSKNASKVANKGKSKS 141
          |||||||||||||||||||||
Db  121 ALWEPKSKNASKVANKGKSKS 141
```

RESULT 2
```
ID   ADES9764 standard; protein; 141 AA.
XX
AC   ADES9764;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Human Protein NP_006324, SEQ ID NO 5660.
XX
KW   Human; pain; neuronal tissue; gene therapy; chronic constriction injury; CCI;
KW   spinal segmental nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
XX
PN   WO020031016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C, D'urso D, Befort K, Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; NP_006324.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
```

```
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 141 AA;
```

```
Query Match          100.0%;  Score 715;  DB 7;  Length 141;
Best Local Similarity 100.0%;  Pred. No. 1.9e-66;
Matches  141;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITDKKKAGKLDRGAASRFVKN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITDKKKAGKLDRGAASRFVKN 120

Qy  121 ALWEPKSKNASKVANKGKSKS 141
          |||||||||||||||||||||
Db  121 ALWEPKSKNASKVANKGKSKS 141
```

RESULT 3
```
ID   AAO15405 standard; protein; 141 AA.
XX
AC   AAO15405;
XX
DT   27-SEP-2002 (first entry)
XX
DE   Mammalian nuclear receptor cofactor CF7 protein.
XX
KW   Mammalian; nuclear receptor cofactor; CF7; CF8; metabolism regulation;
KW   cell homeostasis; cell proliferation; differentiation;
KW   pathological cellular aberration; cellular defence mechanism.
XX
OS   Mammalia.
XX
XX
PN   WO200242322-A2.
XX
PD   30-MAY-2002.
XX
PF   21-NOV-2001; 2001WO-EP013548.
XX
PR   21-NOV-2000; 2000EP-00125524.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Jackson D, Casari G, Suckow J;
XX
DR   WPI; 2002-566559/60.
DR   N-PSDB; AAL43973.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.
XX
PS   Claim 12; Fig 3; 68pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of two
CC   mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC   CF8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CF7
CC   and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC   involved in cellular functions, such as: regulation of metabolism and
```

CC cell homeostasis, cell proliferation and differentiation, pathological
CC cellular aberrations, or cellular defence mechanisms. The present amino
CC acid sequence represents the mammalian nuclear receptor cofactor CF7
CC protein
XX
SQ Sequence 141 AA;

Qy Query Match 93.8%; Score 671; DB 5; Length 141;
Best Local Similarity 93.6%;
Matches 132; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```
Qy    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          ||||||||||||||:|||||:||||:||||||:|||||||||||||||||||||||||||
Db    1 MAAEEINEDYPVEIHDYLSAFANSIDAVDEMLKNMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVRN 120

Qy  121 ALWEPKSKNASKVANKGKSKS 141
          ||||||::|| ||||:|||||
Db  121 ALWEPKPKNTSKVAHGKSKS 141
```

RESULT 4
ADI28083
ID ADI28083 standard; protein; 140 AA.
XX
AC ADI28083;
XX
DT 22-APR-2004 (first entry)
XX
DE Human nuclear receptor cofactor CF6 protein.
XX
KW nuclear receptor cofactor; CF6; cellular function inhibition;
KW metabolic disorder; immunological indication; hormonal dysfunction;
KW neurosystemic disease.
XX
OS Homo sapiens.
XX
XX
PN WO20022472B-A2.
XX
PD 28-MAR-2002.
XX
PF 17-SEP-2001; 2001WO-EP010744.
XX
PR 22-SEP-2000; 2000EP-00120722.
XX
PA (LION-) LION BIOSCIENCE AG.
XX
PI Casari G, Jackson D;
XX
DR WPI; 2002-383179/41.
DR N-PSDB; ADI28090.
XX
PS Claim 12; SEQ ID NO 3; 97pp; English.
XX
CC The invention relates to an isolated polynucleotide encoding a nuclear
CC receptor cofactor, also known as CF6. The polynucleotide or encoded
CC protein is useful for construction of multiple nuclear receptor cofactor
CC specific sequence alignments, preferably for the construction of protein
CC sequence alignments. The protein is useful for screening agents capable
CC of inhibiting the cellular function of the cofactor CF6. The
CC polynucleotide is useful for making vectors and for transforming cells,
CC both of which are ultimately useful for production of the CF6 protein.
CC They are also useful as scientific research tools for developing nucleic
CC acid probes for determining expression levels of the cofactor gene, e.g.
CC to identify diseased or otherwise abnormal states. They are particularly
CC useful for diagnostic purposes to e.g. identify deleted or mutant CF6

CC genes; or their measure expression. They are useful for developing
CC analytical tools such as antisense oligonucleotide for selectively
CC inhibiting expression of the cofactor gene to determine physiological
CC responses. The protein is useful for screening drugs for agonist and
CC antagonist activity, and therefore, for screening for drugs useful in
CC regulating physiological responses associated with the cofactors such as
CC metabolic disorders, immunological indications, hormonal dysfunction,
CC neurosystemic diseases. The proteins are also useful for developing
CC antibodies for detection of the proteins. The polynucleotide can be used
CC to design primers for a polymerase chain reaction and are also used to
CC model the three-dimensional structure of the protein. (Note: this sequence
CC represents the CF6 protein. This sequence appears to be a fragment
CC of the CF6 protein as the sequence differs from the CF6 protein sequence
CC given in Fig 3 (also designated as SEQ ID NO: 2) of the specification).
XX
SQ Sequence 140 AA;

Qy Query Match 93.3%; Score 667; DB 5; Length 140;
Best Local Similarity 93.6%;
Matches 131; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```
Qy    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          ||||||||||||||:|||||:||||:||||||:|||||||||||||||||||||||||||
Db    1 MAAEEINEDYPVEIHDYLSAFANSIDAVDEMLKNMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVRN 120

Qy  121 ALWEPKSKNASKVANKGKSK 140
          ||||||::|| ||||:||||
Db  121 ALWEPKPKNTSKVAHGKSK 140
```

RESULT 5
ADE59762
ID ADE59762 standard; protein; 141 AA.
XX
AC ADE59762;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rat Protein AAH05436, SEQ ID NO 5658.
XX
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; AAH05436.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC New composition comprising two or more isolated polypeptides, useful for
CC preparing a medicament for treating pain in an animal.
CC The invention discloses a composition comprising two or more isolated rat

CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the vector, a host cell
CC   comprising the vector, a host cell
CC   which is differentially regulated in an animal subjected to pain, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain, a method
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for identifying a compound or small molecule that
CC   method for identifying a compound or small molecule that regulates the
CC   method for producing a pharmaceutical composition, a
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   specification, a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide is useful in treating
CC   modulates its activity in an animal (e.g. gene
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic construction
CC   injury (CCI) in an animal (e.g. gene
CC   therapy). The polynucleotide or the compound useful in treating
CC   pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence
CC   presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 141 AA;

Query Match   Best Local Similarity   91.5%;
                        Score 654; DB 7; Length 141;
Matches 127; Conservative   8; Pred. No. 4.5e-60;
                        Mismatches   5; Indels   0; Gaps   0;

Qy   1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
         ||||| |||||||| ::|:: ::|| ||:|||||||:|||||||||||||||||||||||:
Db   1 MAGEMNEDYPVEIHESLITALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120
         |||||||||||||||||||||:|||||||||||||||||||||||||:||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKN 120

Qy 121 ALWEPKSKNASKVANKGKSK 140
         |||||:||:::|||||||||
Db 121 ALWEPKAKSTPKVANKGKSK 140

RESULT 6
ID   AAY51025 standard; protein; 141 AA.
AC   AAY51025;
XX
DT   17-MAR-2000 (first entry)
XX
DE   Murine C1D protein.
XX
KW   C1D; murine; apoptosis; tumour; gene therapy; treatment.
XX
OS   Mus sp.
XX
PN   DE19824811-A1.
XX
PD   09-DEC-1999.
XX
PF   03-JUN-1998; 98DE-01024811.
PR   03-JUN-1998; 98DE-01024811.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Rothbarth K, Stammer H, Werner D;
XX
DR   WPI; 2000-063506/06.

DR   N-PSDB; AAZ43928.
XX
PT   Inducing apoptosis by overexpressing the C1D gene, particularly for
PT   treating tumors.
XX
PS   Claim 3; Fig 2; 10pp; German.
XX
CC   This invention describes a novel method for inducing apoptosis which
CC   comprises overexpressing the C1D gene (I). The method is particularly
CC   used to treat tumors and can also be used in gene therapy. The method has
CC   no side effects on normal cells (contrast known methods of inducing
CC   apoptosis such as cytotoxins and radiation), and may be effective on
CC   cells resistant to conventional treatments. Overexpression of (I) is
CC   sufficient itself to induce apoptosis but the effect may be increased
CC   when used in combination with other anti-tumor methods. When cells
CC   transfected with (I) undergo apoptosis, they release factors that kill
CC   neighboring, non-transfected cells (bystander effect). This sequence
CC   represents the murine C1D protein described in the method of the
CC   invention.
XX
SQ   Sequence 141 AA;

Query Match   Best Local Similarity   90.3%;
                        Score 646; DB 3; Length 141;
Matches 126; Conservative   7; Pred. No. 3.1e-59;
                        Mismatches   7; Indels   0; Gaps   0;

Qy   1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
         ||||| |||||||| ::|:: ::|| ||:|||||||:|||||||||||||||||||||||
Db   1 MAGEMNEDYPVEIHESLITALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120
         ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK 120

Qy 121 ALWEPKRKSTPKVANKGKSK 140
         ||||||:|: |||||||||
Db 121 ALWEPKAKSTPKVANKGKSK 140

RESULT 7
ID   AAO15406 standard; protein; 140 AA.
AC   AAO15406;
XX
DT   27-SEP-2002 (first entry)
XX
DE   Mammalian nuclear receptor cofactor CF8 protein.
XX
KW   Mammalian; nuclear receptor cofactor; CF7; CF8; metabolism regulation;
KW   cell homeostasis; cell proliferation; differentiation;
KW   pathological cellular aberration; cellular defence mechanism.
XX
OS   Mammalia.
XX
PN   WO200242322-A2.
XX
PD   30-MAY-2002.
XX
PF   21-NOV-2001; 2001WO-EP013548.
PR   21-NOV-2000; 2000EP-00125524.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Jackson D, Casari G, Suckow J;
XX
DR   WPI; 2002-566559/60.
DR   N-PSDB; AAL43975.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.

XX
PS Claim 12; Fig 3; 68pp; English.
XX
CC The invention comprises the amino acid and coding sequences of two
CC mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC CF8 protein sequences of the invention are useful for screening agents
CC that are capable of inhibiting the cellular function of cofactor CF7
CC and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC involved in cellular functions, such as: regulation of metabolism and
CC cell homeostasis, cell proliferation and differentiation, pathological
CC cellular aberrations, or cellular defence mechanisms. The present amino
CC acid sequence represents the mammalian nuclear receptor cofactor CF8
CC protein
XX
SQ Sequence 140 AA;

Qy Query Match 90.1%; Score 644; DB 5; Length 140;
Best Local Similarity 91.4%; Pred. No. 5e-59;
Matches 128; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy   2 AGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSAY 61
          |||||||||||||| |||||| ||| ||||||| |||||||||||||||||||||||||||
Db   1 AAEEINEDYPVEIHDYLSAFANSIDAVDEMLKNMMSVSRNELLQKLDPLEQAKVDLVSAY 60

Qy   62 TLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITDKKKAGKLDRGAASRFVKRNA 121
          |||||| |||||||||||| |||||||||||||||||| ||| |||||||||||||| ||
Db   61 ILNSMFWAYLATQGVNPKEHSVKQELERIRVYMRVKEIADKKXAGKLDRGAASRFVRNA 120

Qy 122 LWEPKSKNASKVANKGKSKS 141
          ||||| ||||||| ||||||
Db 121 LWEPKPKNASKVAYKGKSKS 140

RESULT 8
AAG01231
ID AAG01231 standard; protein; 102 AA.
XX
AC AAG01231;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 5312.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-00200610.
XX
PR 26-FEB-1999; 99US-0122487P.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR N-PSDB; AAC01237.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS Claim 13; SEQ ID NO 5312; 71pp + Sequence listing; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number of
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated

CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC sequences derived from the 5' ends of mRNAs and even in those cases where
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors
XX
SQ Sequence 102 AA;

Qy Query Match 72.6%; Score 519; DB 3; Length 102;
Best Local Similarity 100.0%; Pred. No. 3.7e-46;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy 61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITD 102
         |||||||||||||||||||||||||||||||||||||||||
Db 61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITD 102

RESULT 9
AB054730
ID AB054730 standard; protein; 54 AA.
XX
AC AB054730;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human genome derived single exon protein #964.
XX
KW Human; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX
OS Homo sapiens.
XX
XX
PN US2003194704-A1.
XX
PD 16-OCT-2003.
XX
PF 03-APR-2002; 2002US-00029386.
XX
PR 03-APR-2002; 2002US-00029386.
XX
PA (PENN/) PENN S G.
PA (RANK/) RANK D R.
PA (HANZ/) HANZEL D K.
XX
PI Penn SG, Rank DR, Hanzel DK;
XX
DR WPI; 2004-119264/12.
XX
PT New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing event, for assessing genomic alterations or as tools for
PT surveying tissues.
XX
PS Claim 45; SEQ ID NO 28364; 80pp; English.
XX
CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately

CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above- mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subscription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterising
CC alternative splicing events, in detecting and characterising gross
CC alterations in the genomic locus, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe protein of the invention. The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ   Sequence 54 AA;

Qy   88 ERIRVYMNRVKEITDKKKAGKLDRGAASRFVKNALWEPKSKVASKVANKGKSKS 141
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ERIRVYMNRVKEITDKKKAGKLDRGAASRFVKNALWEPKPKVASKVANKGKSKS 54

Qy   Query Match   37.2%;
     Best Local Similarity   98.1%;
     Matches   53; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

RESULT 10
AAG49051
ID   AAG49051 standard; protein; 217 AA.
XX
AC   AAG49051;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62013.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.

XX   Score 266; DB 8; Length 54;
       Pred. No. 4.2e-20;

PR   28-APR-1999; 99US-0131449P.
PR   30-APR-1999; 99US-0120408P.
PR   30-APR-1999; 99US-0132407P.
PR   04-MAY-1999; 99US-0124484P.
PR   05-MAY-1999; 99US-0124485P.
PR   06-MAY-1999; 99US-0124486P.
PR   06-MAY-1999; 99US-0132487P.
PR   07-MAY-1999; 99US-0128263P.
PR   11-MAY-1999; 99US-0134256P.
PR   14-MAY-1999; 99US-0134218P.
PR   14-MAY-1999; 99US-0134221P.
PR   14-MAY-1999; 99US-0134768P.
PR   18-MAY-1999; 99US-0134769P.
PR   19-MAY-1999; 99US-0134941P.
PR   20-MAY-1999; 99US-0134370P.
PR   21-MAY-1999; 99US-0135353P.
PR   24-MAY-1999; 99US-0135629P.
PR   25-MAY-1999; 99US-0136021P.
PR   27-MAY-1999; 99US-0136392P.
PR   28-MAY-1999; 99US-0136393P.
PR   01-JUN-1999; 99US-0136782P.
PR   03-JUN-1999; 99US-0137222P.
PR   04-JUN-1999; 99US-0137528P.
PR   07-JUN-1999; 99US-0137502P.
PR   07-JUN-1999; 99US-0137724P.
PR   08-JUN-1999; 99US-0138094P.
PR   10-JUN-1999; 99US-0138540P.
PR   10-JUN-1999; 99US-0138847P.
PR   14-JUN-1999; 99US-0139119P.
PR   16-JUN-1999; 99US-0139460P.
PR   16-JUN-1999; 99US-0139461P.
PR   18-JUN-1999; 99US-0139452P.
PR   17-JUN-1999; 99US-0139453P.
PR   18-JUN-1999; 99US-0139454P.
PR   18-JUN-1999; 99US-0139455P.
PR   18-JUN-1999; 99US-0139456P.
PR   18-JUN-1999; 99US-0139457P.
PR   18-JUN-1999; 99US-0139458P.
PR   18-JUN-1999; 99US-0139459P.
PR   18-JUN-1999; 99US-0139460P.
PR   18-JUN-1999; 99US-0139461P.
PR   18-JUN-1999; 99US-0139462P.
PR   18-JUN-1999; 99US-0139463P.
PR   18-JUN-1999; 99US-0139750P.
PR   21-JUN-1999; 99US-0139817P.
PR   23-JUN-1999; 99US-0138999P.
PR   23-JUN-1999; 99US-0140353P.
PR   24-JUN-1999; 99US-0140354P.
PR   28-JUN-1999; 99US-0140695P.
PR   29-JUN-1999; 99US-0140699P.
PR   30-JUN-1999; 99US-0140823P.
PR   01-JUL-1999; 99US-0140991P.
PR   01-JUL-1999; 99US-0141287P.
PR   01-JUL-1999; 99US-0141842P.
PR   02-JUL-1999; 99US-0142154P.
PR   06-JUL-1999; 99US-0142055P.
PR   08-JUL-1999; 99US-0142390P.
PR   09-JUL-1999; 99US-0142803P.
PR   09-JUL-1999; 99US-0142920P.
PR   12-JUL-1999; 99US-0142977P.
PR   13-JUL-1999; 99US-0143542P.
PR   14-JUL-1999; 99US-0143624P.
PR   15-JUL-1999; 99US-0144005P.
PR   16-JUL-1999; 99US-0144085P.
PR   16-JUL-1999; 99US-0144086P.
PR   19-JUL-1999; 99US-0143325P.
PR   19-JUL-1999; 99US-0144331P.
PR   19-JUL-1999; 99US-0144332P.
PR   19-JUL-1999; 99US-0144333P.
PR   19-JUL-1999; 99US-0144334P.
PR   20-JUL-1999; 99US-0143352P.
PR   20-JUL-1999; 99US-0144632P.
PR   20-JUL-1999; 99US-0148884P.

PR   25-FEB-1999; 99US-0121825P.
PR   05-MAR-1999; 99US-0123180P.
PR   09-MAR-1999; 99US-0123548P.
PR   23-MAR-1999; 99US-0125788P.
PR   25-MAR-1999; 99US-0126264P.
PR   29-MAR-1999; 99US-0126785P.
PR   01-APR-1999; 99US-0127462P.
PR   06-APR-1999; 99US-0128234P.
PR   08-APR-1999; 99US-0128714P.
PR   16-APR-1999; 99US-0129845P.
PR   19-APR-1999; 99US-0130077P.
PR   21-APR-1999; 99US-0130449P.
PR   23-APR-1999; 99US-0130510P.
PR   23-APR-1999; 99US-0130891P.

PR 21-JUL-1999; 99US-0148814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 27-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147302P.
PR 04-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 05-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 06-AUG-1999; 99US-0147935P.
PR 09-AUG-1999; 99US-0148171P.
PR 10-AUG-1999; 99US-0148341P.
PR 11-AUG-1999; 99US-0148565P.
PR 12-AUG-1999; 99US-0149360P.
PR 13-AUG-1999; 99US-0149175P.
PR 13-AUG-1999; 99US-0149426P.
PR 16-AUG-1999; 99US-0149722P.
PR 17-AUG-1999; 99US-0149723P.
PR 18-AUG-1999; 99US-0149929P.
PR 20-AUG-1999; 99US-0149902P.
PR 20-AUG-1999; 99US-0149930P.
PR 20-AUG-1999; 99US-0150566P.
PR 23-AUG-1999; 99US-0150804P.
PR 23-AUG-1999; 99US-0151065P.
PR 25-AUG-1999; 99US-0151066P.
PR 26-AUG-1999; 99US-0151080P.
PR 27-AUG-1999; 99US-0151303P.
PR 27-AUG-1999; 99US-0151438P.
PR 30-AUG-1999; 99US-0151930P.
PR 31-AUG-1999; 99US-0152363P.
PR 01-SEP-1999; 99US-0152548P.
PR 07-SEP-1999; 99US-0153070P.
PR 10-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 15-SEP-1999; 99US-0154039P.
PR 16-SEP-1999; 99US-0154779P.
PR 20-SEP-1999; 99US-0155139P.
PR 20-SEP-1999; 99US-0155486P.
PR 22-SEP-1999; 99US-0155659P.
PR 23-SEP-1999; 99US-0155496P.
PR 24-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 29-SEP-1999; 99US-0157117P.
PR 04-OCT-1999; 99US-0157753P.
PR 05-OCT-1999; 99US-0157865P.
PR 06-OCT-1999; 99US-0158029P.
PR 07-OCT-1999; 99US-0158232P.
PR 08-OCT-1999; 99US-0158369P.
PR 12-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159299P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159377P.
PR 14-OCT-1999; 99US-0159638P.

PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 22-OCT-1999; 99US-0161359P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 28-OCT-1999; 99US-0161406P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match   21.0%;   Score 150.5;   DB 3;   Length 217;
Best Local Similarity   27.6%;   Pred. No. 2.8e-07;
Matches 37;   Conservative 35;   Mismatches 47;   Indels 15;   Gaps 2;

```
Qy     22 ENSIGAVDE----------MLKTMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWYL 71
             |::| ||::  |:::| |:| ||::|||::|:||:::::
Db     18 ESAIEAVNQTLAYLKELKPQLEQMLTLAEPEVLAAMQPLQRAKTMHLLAEATTTLYELRL 77

Qy     72 ATQGVNPKEHPVKQELERIRVTYMNRVKEITDKKK-------AGKLDRGAASRFVKNALWEPK 126
             |:|:| | || |||||::|||||| |:|: ||:|||| ::|::|
Db     78 RCTGVDPDDHRVKSEIERINVYREKPQKCVDQSKGPLRPTTVLNRQAATRFIEHSLPDLT 137

Qy    127 SKNASKVANKGKSK 140
              :|:|||::|:
Db    138 STQKQSIRDLSKGE 151
```

RESULT 11
AAG24127
ID   AAG24127 standard; protein; 217 AA.
XX
AC   AAG24127;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27683.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   BP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000BP-00301439.

PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0125548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130499P.
PR 23-APR-1999; 99US-0130510P.

```
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134089P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 21-MAY-1999; 99US-0135124P.
PR 24-MAY-1999; 99US-0135353P.
PR 25-MAY-1999; 99US-0135629P.
PR 27-MAY-1999; 99US-0136021P.
PR 28-MAY-1999; 99US-0136392P.
PR 01-JUN-1999; 99US-0136782P.
PR 03-JUN-1999; 99US-0137222P.
PR 04-JUN-1999; 99US-0137528P.
PR 07-JUN-1999; 99US-0137502P.
PR 08-JUN-1999; 99US-0137724P.
PR 10-JUN-1999; 99US-0138094P.
PR 14-JUN-1999; 99US-0138540P.
PR 16-JUN-1999; 99US-0138847P.
PR 17-JUN-1999; 99US-0139119P.
PR 18-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 22-JUN-1999; 99US-0139817P.
PR 23-JUN-1999; 99US-0139899P.
PR 24-JUN-1999; 99US-0140353P.
PR 28-JUN-1999; 99US-0140354P.
PR 29-JUN-1999; 99US-0140695P.
PR 30-JUN-1999; 99US-0140823P.
PR 01-JUL-1999; 99US-0140991P.
PR 01-JUL-1999; 99US-0141287P.
PR 02-JUL-1999; 99US-0141842P.
PR 06-JUL-1999; 99US-0142154P.
PR 08-JUL-1999; 99US-0142055P.
PR 09-JUL-1999; 99US-0142290P.
PR 12-JUL-1999; 99US-0142803P.
PR 13-JUL-1999; 99US-0142920P.
PR 14-JUL-1999; 99US-0142977P.
PR 15-JUL-1999; 99US-0143542P.
PR 16-JUL-1999; 99US-0143624P.
PR 16-JUL-1999; 99US-0144005P.
PR 19-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0174493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150804P.
PR 27-AUG-1999; 99US-0150651P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151438P.
PR 07-SEP-1999; 99US-0151930P.
PR 10-SEP-1999; 99US-0152363P.
PR 13-SEP-1999; 99US-0153070P.
PR 15-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 20-SEP-1999; 99US-0154039P.
PR 22-SEP-1999; 99US-0154779P.
PR 23-SEP-1999; 99US-0155139P.
PR 24-SEP-1999; 99US-0155486P.
PR 28-SEP-1999; 99US-0155596P.
PR 29-SEP-1999; 99US-0156458P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
```

```
RESULT 12
AAG49050
ID   AAG49050 standard; protein; 253 AA.
XX
AC   AAG49050;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62012.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
XX
PN   EP103405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999; 99US-0121825P.
PR   05-MAR-1999; 99US-0123180P.
PR   09-MAR-1999; 99US-0123548P.
PR   23-MAR-1999; 99US-0125788P.
PR   25-MAR-1999; 99US-0126264P.
PR   29-MAR-1999; 99US-0126785P.
PR   01-APR-1999; 99US-0127462P.
PR   06-APR-1999; 99US-0128234P.
PR   08-APR-1999; 99US-0128714P.
PR   16-APR-1999; 99US-0129845P.
PR   19-APR-1999; 99US-0130077P.
PR   21-APR-1999; 99US-0130449P.
PR   23-APR-1999; 99US-0130510P.
PR   23-APR-1999; 99US-0130891P.
PR   28-APR-1999; 99US-0131449P.
PR   30-APR-1999; 99US-0132048P.
PR   30-APR-1999; 99US-0132407P.
PR   04-MAY-1999; 99US-0132484P.
PR   05-MAY-1999; 99US-0132486P.
PR   06-MAY-1999; 99US-0132487P.
PR   07-MAY-1999; 99US-0132863P.
PR   11-MAY-1999; 99US-0134256P.
PR   14-MAY-1999; 99US-0134218P.
PR   14-MAY-1999; 99US-0134219P.
PR   14-MAY-1999; 99US-0134221P.
PR   17-MAY-1999; 99US-0134768P.
PR   19-MAY-1999; 99US-0134370P.
PR   20-MAY-1999; 99US-0135124P.
PR   21-MAY-1999; 99US-0135353P.
PR   24-MAY-1999; 99US-0135299P.
PR   25-MAY-1999; 99US-0136021P.
PR   27-MAY-1999; 99US-0136782P.
PR   28-MAY-1999; 99US-0136782P.
PR   01-JUN-1999; 99US-0137222P.
PR   03-JUN-1999; 99US-0137528P.
PR   04-JUN-1999; 99US-0137502P.
PR   07-JUN-1999; 99US-0137724P.
PR   08-JUN-1999; 99US-0138094P.
PR   10-JUN-1999; 99US-0138540P.
PR   10-JUN-1999; 99US-0138847P.
PR   14-JUN-1999; 99US-0139119P.
PR   16-JUN-1999; 99US-0139452P.
PR   16-JUN-1999; 99US-0139453P.
PR   17-JUN-1999; 99US-0139492P.
PR   18-JUN-1999; 99US-0139454P.
PR   18-JUN-1999; 99US-0139455P.
PR   18-JUN-1999; 99US-0139456P.
PR   18-JUN-1999; 99US-0139458P.
PR   18-JUN-1999; 99US-0139459P.
PR   18-JUN-1999; 99US-0139460P.
PR   18-JUN-1999; 99US-0139461P.
PR   18-JUN-1999; 99US-0139462P.
PR   18-JUN-1999; 99US-0139463P.
PR   18-JUN-1999; 99US-0139750P.
PR   21-JUN-1999; 99US-0139817P.
PR   22-JUN-1999; 99US-0139899P.
PR   23-JUN-1999; 99US-0140353P.
PR   23-JUN-1999; 99US-0140354P.
PR   24-JUN-1999; 99US-0140695P.
PR   28-JUN-1999; 99US-0140823P.
PR   29-JUN-1999; 99US-0140991P.
PR   30-JUN-1999; 99US-0141287P.
PR   01-JUL-1999; 99US-0141842P.
PR   01-JUL-1999; 99US-0142154P.
PR   02-JUL-1999; 99US-0142155P.
PR   06-JUL-1999; 99US-0142390P.
PR   08-JUL-1999; 99US-0142803P.
PR   09-JUL-1999; 99US-0142920P.
PR   12-JUL-1999; 99US-0142977P.
PR   13-JUL-1999; 99US-0143542P.
PR   14-JUL-1999; 99US-0143624P.
PR   15-JUL-1999; 99US-0144005P.
PR   16-JUL-1999; 99US-0144086P.
PR   16-JUL-1999; 99US-0144325P.
PR   19-JUL-1999; 99US-0144331P.
PR   19-JUL-1999; 99US-0144332P.
PR   19-JUL-1999; 99US-0144333P.
PR   19-JUL-1999; 99US-0144334P.
PR   19-JUL-1999; 99US-0144335P.
PR   20-JUL-1999; 99US-0144352P.
PR   14-OCT-1999; 99US-0159638P.
PR   18-OCT-1999; 99US-0159584P.
PR   21-OCT-1999; 99US-0160741P.
PR   21-OCT-1999; 99US-0160767P.
PR   21-OCT-1999; 99US-0160768P.
PR   21-OCT-1999; 99US-0160770P.
PR   21-OCT-1999; 99US-0160814P.
PR   21-OCT-1999; 99US-0160815P.
PR   22-OCT-1999; 99US-0160980P.
PR   22-OCT-1999; 99US-0160981P.
PR   22-OCT-1999; 99US-0160989P.
PR   25-OCT-1999; 99US-0161404P.
PR   25-OCT-1999; 99US-0161405P.
PR   25-OCT-1999; 99US-0161406P.
PR   26-OCT-1999; 99US-0161360P.
PR   26-OCT-1999; 99US-0161361P.
PR   26-OCT-1999; 99US-0161359P.
PR   28-OCT-1999; 99US-0161920P.
PR   28-OCT-1999; 99US-0161992P.
PR   28-OCT-1999; 99US-0161993P.
PR   29-OCT-1999; 99US-0162142P.

Query Match             21.0%;  Score 150.5; DB 3;  Length 217;
Best Local Similarity   27.6%;  Pred. No. 2.8e-07;
Matches   37;  Conservative  35;  Mismatches  47;  Indels  15;  Gaps   2;

Qy   22 ENSIGAVDE----------MLKTWMSVSRNELLQKLDPLEQAKYDLVSAYTLNSMFWVYL 71
            :|  |||:              |: :::|| | |:  |: | ::| : :| :: |
Db   18 ESAIEAVNQTLAVLKELKPQLEQMLTLAEPEVLAAMQPLQRAKTMHLLAEATTLYEBLRL 77

Qy   72 ATQGVNPKEHPVKQELERIRVYMNRVKEITDKKK----------AGKLDRGAASRFVKNALWEPK 126
            : |  | |::||| ||ERI.VY :|.| :          ::  |  :  :||:|  : ::|
Db   78 RCTGVDPDDHRVKSEIERINVYREKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSLPDLT 137

Qy  127 SKNASKVANKGKSK 140
            | : | :::|| :
Db  138 STQKQSIRDLSKGE 151
```

PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145008P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147202P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150584P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.

PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 26-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 28-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 29-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match   Best Local Similarity   21.0%;   Score 150.5;   DB 3;   Length 253;
Best Local Similarity   21.0%;
Matches   37;   Conservative   35;   Mismatches   47;   Indels   15;   Gaps   2;

Qy   22 ENSIGAVDB----------MLKTMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYL 71
             :| |:::| |:| |:::|||| ::| ::| | ||:|
Db   54 ESAIEAVNQTLAYLKELKPQLEQMLTLAEPEVLAAMQPLQRAKTMHLLAEATTTLYELRL 113

Qy   72 ATQGVNPKEHPVKQELERIRVYMNRVKEITDKKK------AGKLDRGAASRFVKNALWEPK 126
         |:| | ::||:| :|||:||| :::|: ||| || | :|||
Db 114 RCTGVDPDDHRVKSEIERINVYREKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSLPDLT 173

Qy 127 SKNASKVANKGKSK 140
            : |:
Db 174 STQKQSIRDLSKGE 187

RESULT 13
AAG24126
ID   AAG24126 standard; protein; 256 AA.
AC   AAG24126;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27682.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.

```
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 21-MAY-1999; 99US-0135124P.
PR 24-MAY-1999; 99US-0135353P.
PR 25-MAY-1999; 99US-0355629P.
PR 27-MAY-1999; 99US-0356021P.
PR 28-MAY-1999; 99US-0136392P.
PR 01-JUN-1999; 99US-0136782P.
PR 03-JUN-1999; 99US-0137222P.
PR 04-JUN-1999; 99US-0137528P.
PR 07-JUN-1999; 99US-0137502P.
PR 08-JUN-1999; 99US-0137724P.
PR 10-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0138847P.
PR 16-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 22-JUN-1999; 99US-0139817P.
PR 23-JUN-1999; 99US-0139899P.
PR 24-JUN-1999; 99US-0140353P.
PR 28-JUN-1999; 99US-0140354P.
PR 29-JUN-1999; 99US-0140695P.
PR 30-JUN-1999; 99US-0140823P.
PR 01-JUL-1999; 99US-0140991P.
PR 01-JUL-1999; 99US-0141287P.
PR 02-JUL-1999; 99US-0141842P.
PR 06-JUL-1999; 99US-0142154P.
PR 08-JUL-1999; 99US-0142055P.
PR 09-JUL-1999; 99US-0142390P.
PR 12-JUL-1999; 99US-0142803P.
PR 13-JUL-1999; 99US-0142920P.
PR 14-JUL-1999; 99US-0142977P.
PR 15-JUL-1999; 99US-0143542P.
PR 16-JUL-1999; 99US-0143624P.
PR 16-JUL-1999; 99US-0144005P.
PR 19-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0143152P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0148014P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 04-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 05-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 06-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 09-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 10-AUG-1999; 99US-0147493P.
PR 11-AUG-1999; 99US-0147335P.
PR 12-AUG-1999; 99US-0148171P.
PR 13-AUG-1999; 99US-0148319P.
PR 13-AUG-1999; 99US-0148341P.
PR 16-AUG-1999; 99US-0148565P.
PR 17-AUG-1999; 99US-0148684P.
PR 18-AUG-1999; 99US-0148175P.
PR 20-AUG-1999; 99US-0149368P.
PR 20-AUG-1999; 99US-0149175P.
PR 20-AUG-1999; 99US-0149222P.
PR 23-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149929P.
PR 25-AUG-1999; 99US-0149902P.
PR 26-AUG-1999; 99US-0149930P.
PR 27-AUG-1999; 99US-0150566P.
PR 27-AUG-1999; 99US-0150884P.
PR 30-AUG-1999; 99US-0151065P.
PR 31-AUG-1999; 99US-0151066P.
PR 01-SEP-1999; 99US-0151080P.
PR 07-SEP-1999; 99US-0151303P.
PR 10-SEP-1999; 99US-0151438P.
PR 13-SEP-1999; 99US-0151930P.
PR 15-SEP-1999; 99US-0152363P.
PR 16-SEP-1999; 99US-0153758P.
PR 20-SEP-1999; 99US-0153070P.
PR 22-SEP-1999; 99US-0154018P.
PR 24-SEP-1999; 99US-0154039P.
PR 28-SEP-1999; 99US-0154779P.
PR 29-SEP-1999; 99US-0155139P.
PR 04-OCT-1999; 99US-0155486P.
PR 05-OCT-1999; 99US-0155659P.
PR 06-OCT-1999; 99US-0155458P.
PR 07-OCT-1999; 99US-0156596P.
PR 08-OCT-1999; 99US-0157117P.
PR 12-OCT-1999; 99US-0157553P.
PR 13-OCT-1999; 99US-0157865P.
PR 13-OCT-1999; 99US-0158029P.
PR 13-OCT-1999; 99US-0158232P.
PR 14-OCT-1999; 99US-0158369P.
PR 14-OCT-1999; 99US-0158294P.
PR 14-OCT-1999; 99US-0159293P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
```

```
PR   14-OCT-1999;   99US-0159331P.
PR   14-OCT-1999;   99US-0159637P.
PR   18-OCT-1999;   99US-0159628P.
PR   21-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160899P.
PR   22-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   25-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.
```

Query Match   Best Local Similarity 27.6%; Score 150.5; DB 3; Length 256;
Matches 37; Conservative 35; Mismatches 47; Indels 15; Gaps 2;

```
Qy   22  ENSIGAVDE----------MLKTMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYL  71
             :|:  ||::                  |||:    |:  ::    :|   ::|   :|:
Db   57  ESAIEAVNQTLAYLKELKPQLEQMLTLAEPEVLAAMQPLQRAKTMHLLAEATTLYELRL  116

Qy   72  ATQGVNPKEHPVKQELERIRVYMNRVKEITDKKK--------AGKLDRGAASRFVKNALWEPK  126
             |:|   |:|   ||:|||:|| :: |::::         | :||  :|:::  ||:|  :
Db  117  RCTGVDPDDIRVKSEIERINVYYREKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSLPDLT  176

Qy  127  SKNASKVANKGKSK  140
             :|    ::: |:
Db  177  STQKQSIRDLSKGE  190
```

RESULT 14
AAG24128

```
ID   AAG24128 standard; protein; 177 AA.
XX
AC   AAG24128;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27684.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;  2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825P.
PR   05-MAR-1999;  99US-0123180P.
PR   09-MAR-1999;  99US-0123549P.
PR   23-MAR-1999;  99US-0125788P.
PR   25-MAR-1999;  99US-0126264P.
PR   29-MAR-1999;  99US-0126785P.
PR   01-APR-1999;  99US-0127462P.
PR   06-APR-1999;  99US-0127678P.
PR   08-APR-1999;  99US-0128234P.
PR   16-APR-1999;  99US-0129845P.
PR   19-APR-1999;  99US-0130077P.
PR   21-APR-1999;  99US-0130449P.
PR   23-APR-1999;  99US-0130510P.
PR   28-APR-1999;  99US-0130891P.
PR   28-APR-1999;  99US-0131494P.
PR   30-APR-1999;  99US-0132048P.
PR   30-APR-1999;  99US-0132407P.
PR   04-MAY-1999;  99US-0132484P.
PR   05-MAY-1999;  99US-0132485P.
PR   06-MAY-1999;  99US-0132486P.
PR   07-MAY-1999;  99US-0132863P.
PR   11-MAY-1999;  99US-0134258P.
PR   14-MAY-1999;  99US-0134219P.
PR   14-MAY-1999;  99US-0134221P.
PR   14-MAY-1999;  99US-0134370P.
PR   18-MAY-1999;  99US-0134768P.
PR   19-MAY-1999;  99US-0134941P.
PR   20-MAY-1999;  99US-0135124P.
PR   21-MAY-1999;  99US-0135353P.
PR   24-MAY-1999;  99US-0135629P.
PR   25-MAY-1999;  99US-0136021P.
PR   27-MAY-1999;  99US-0136792P.
PR   28-MAY-1999;  99US-0136782P.
PR   01-JUN-1999;  99US-0137222P.
PR   03-JUN-1999;  99US-0137502P.
PR   04-JUN-1999;  99US-0137724P.
PR   07-JUN-1999;  99US-0138094P.
PR   08-JUN-1999;  99US-0138540P.
PR   10-JUN-1999;  99US-0138847P.
PR   14-JUN-1999;  99US-0139110P.
PR   16-JUN-1999;  99US-0139452P.
PR   18-JUN-1999;  99US-0139453P.
PR   18-JUN-1999;  99US-0139492P.
PR   18-JUN-1999;  99US-0139454P.
PR   18-JUN-1999;  99US-0139455P.
PR   18-JUN-1999;  99US-0139456P.
PR   18-JUN-1999;  99US-0139457P.
PR   18-JUN-1999;  99US-0139458P.
PR   18-JUN-1999;  99US-0139459P.
PR   18-JUN-1999;  99US-0139460P.
PR   18-JUN-1999;  99US-0139461P.
PR   18-JUN-1999;  99US-0139462P.
PR   18-JUN-1999;  99US-0139463P.
PR   21-JUN-1999;  99US-0139750P.
PR   22-JUN-1999;  99US-0139763P.
PR   22-JUN-1999;  99US-0139817P.
PR   23-JUN-1999;  99US-0139999P.
PR   23-JUN-1999;  99US-0140353P.
PR   24-JUN-1999;  99US-0140354P.
PR   28-JUN-1999;  99US-0140695P.
PR   29-JUN-1999;  99US-0140823P.
PR   30-JUN-1999;  99US-0140991P.
PR   01-JUL-1999;  99US-0141287P.
PR   01-JUL-1999;  99US-0141842P.
PR   02-JUL-1999;  99US-0142154P.
PR   06-JUL-1999;  99US-0142055P.
PR   08-JUL-1999;  99US-0142803P.
PR   09-JUL-1999;  99US-0142390P.
PR   12-JUL-1999;  99US-0142920P.
PR   13-JUL-1999;  99US-0142777P.
PR   14-JUL-1999;  99US-0143542P.
PR   15-JUL-1999;  99US-0143624P.
PR   16-JUL-1999;  99US-0144005P.
PR   16-JUL-1999;  99US-0144005P.
PR   16-JUL-1999;  99US-0140086P.
PR   19-JUL-1999;  99US-0144325P.
PR   19-JUL-1999;  99US-0143331P.
PR   19-JUL-1999;  99US-0143332P.
PR   19-JUL-1999;  99US-0143333P.
PR   19-JUL-1999;  99US-0144334P.
```

RESULT 15

AAG49052

ID   AAG49052 standard; protein; 177 AA.
XX
AC   AAG49052;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62014.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX

Query Match   19.8%; Score 141.5; DB 3; Length 177;
Best Local Similarity   28.8%; Pred. No. 1.9e-06;
Matches   32; Conservative   30; Mismatches   44; Indels   5; Gaps   1;

Qy   35 MMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVVYLIATQGVNPKEHPVKQELERIRVYM 94
             |::|| |:| |:||| | |||||: ||| |: ::|| || | |||||| |
Db   1 MLTLAEPEVLAAAMQPLQRAKTMHLLAEATTTLYELRLRCTGVDPDDHRVKSEIERINVYR 60

Qy   95 NRVKEITDKKK------AGKLDRGAASRFVKNALWBPKSKNASKVANKGKSK 140
            |:|:: |:| |: | | | |: |: :|
Db   61 EKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSLPDLTSTQKQSIRDLSKGE 111

PR   19-JUL-1999;   99US-0144335P.
PR   20-JUL-1999;   99US-0144352P.
PR   20-JUL-1999;   99US-0144632P.
PR   20-JUL-1999;   99US-0144884P.
PR   21-JUL-1999;   99US-0144814P.
PR   21-JUL-1999;   99US-0145086P.
PR   21-JUL-1999;   99US-0145088P.
PR   22-JUL-1999;   99US-0145085P.
PR   22-JUL-1999;   99US-0145087P.
PR   22-JUL-1999;   99US-0145089P.
PR   22-JUL-1999;   99US-0145192P.
PR   23-JUL-1999;   99US-0145145P.
PR   23-JUL-1999;   99US-0145218P.
PR   28-JUL-1999;   99US-0145951P.
PR   27-JUL-1999;   99US-0145919P.
PR   27-JUL-1999;   99US-0145918P.
PR   26-JUL-1999;   99US-0145913P.
PR   23-JUL-1999;   99US-0145224P.
PR   23-JUL-1999;   99US-0152276P.
PR   04-AUG-1999;   99US-0147302P.
PR   04-AUG-1999;   99US-0147204P.
PR   03-AUG-1999;   99US-0147038P.
PR   02-AUG-1999;   99US-0146389P.
PR   02-AUG-1999;   99US-0146388P.
PR   02-AUG-1999;   99US-0146388P.
PR   11-AUG-1999;   99US-0148319P.
PR   10-AUG-1999;   99US-0148171P.
PR   09-AUG-1999;   99US-0147915P.
PR   09-AUG-1999;   99US-0147493P.
PR   06-AUG-1999;   99US-0147416P.
PR   05-AUG-1999;   99US-0147303P.
PR   05-AUG-1999;   99US-0147260P.
PR   05-AUG-1999;   99US-0147192P.
PR   17-AUG-1999;   99US-0149175P.
PR   16-AUG-1999;   99US-0149368P.
PR   13-AUG-1999;   99US-0148684P.
PR   13-AUG-1999;   99US-0148565P.
PR   12-AUG-1999;   99US-0148341P.
PR   18-AUG-1999;   99US-0149426P.
PR   20-AUG-1999;   99US-0149722P.
PR   20-AUG-1999;   99US-0149723P.
PR   20-AUG-1999;   99US-0149929P.
PR   23-AUG-1999;   99US-0149902P.
PR   23-AUG-1999;   99US-0149930P.
PR   25-AUG-1999;   99US-0150566P.
PR   26-AUG-1999;   99US-0150884P.
PR   27-AUG-1999;   99US-0151065P.
PR   27-AUG-1999;   99US-0151066P.
PR   27-AUG-1999;   99US-0151080P.
PR   30-AUG-1999;   99US-0151303P.
PR   31-AUG-1999;   99US-0151438P.
PR   01-SEP-1999;   99US-0151930P.
PR   07-SEP-1999;   99US-0152363P.
PR   10-SEP-1999;   99US-0153070P.
PR   13-SEP-1999;   99US-0153758P.
PR   15-SEP-1999;   99US-0154018P.
PR   16-SEP-1999;   99US-0154039P.
PR   20-SEP-1999;   99US-0154779P.
PR   22-SEP-1999;   99US-0155139P.
PR   23-SEP-1999;   99US-0155486P.
PR   24-SEP-1999;   99US-0155659P.
PR   28-SEP-1999;   99US-0156458P.
PR   29-SEP-1999;   99US-0156596P.
PR   04-OCT-1999;   99US-0157117P.
PR   05-OCT-1999;   99US-0157753P.
PR   06-OCT-1999;   99US-0157865P.
PR   07-OCT-1999;   99US-0158029P.
PR   08-OCT-1999;   99US-0158232P.
PR   12-OCT-1999;   99US-0158369P.
PR   13-OCT-1999;   99US-0158923P.
PR   13-OCT-1999;   99US-0159293P.
PR   13-OCT-1999;   99US-0159294P.
PR   14-OCT-1999;   99US-0159329P.

PR   14-OCT-1999;   99US-0159330P.
PR   14-OCT-1999;   99US-0159331P.
PR   14-OCT-1999;   99US-0159637P.
PR   14-OCT-1999;   99US-0159638P.
PR   18-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.

PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.

```
PR  23-APR-1999;   99US-0130891P,
PR  28-APR-1999;   99US-0131449P,
PR  30-APR-1999;   99US-0132048P,
PR  30-APR-1999;   99US-0132407P,
PR  04-APR-1999;   99US-0132484P,
PR  05-MAY-1999;   99US-0132485P,
PR  06-MAY-1999;   99US-0132486P,
PR  06-MAY-1999;   99US-0132487P,
PR  07-MAY-1999;   99US-0132863P,
PR  11-MAY-1999;   99US-0134256P,
PR  14-MAY-1999;   99US-0134256P,
PR  14-MAY-1999;   99US-0134218P,
PR  14-MAY-1999;   99US-0134219P,
PR  14-MAY-1999;   99US-0134221P,
PR  19-MAY-1999;   99US-0134370P,
PR  20-MAY-1999;   99US-0134768P,
PR  21-MAY-1999;   99US-0134941P,
PR  24-MAY-1999;   99US-0135124P,
PR  25-MAY-1999;   99US-0135353P,
PR  27-MAY-1999;   99US-0135629P,
PR  28-MAY-1999;   99US-0136021P,
PR  01-JUN-1999;   99US-0136392P,
PR  03-JUN-1999;   99US-0136782P,
PR  04-JUN-1999;   99US-0137222P,
PR  07-JUN-1999;   99US-0137528P,
PR  08-JUN-1999;   99US-0137502P,
PR  10-JUN-1999;   99US-0137724P,
PR  10-JUN-1999;   99US-0138094P,
PR  14-JUN-1999;   99US-0138540P,
PR  16-JUN-1999;   99US-0138847P,
PR  17-JUN-1999;   99US-0139119P,
PR  18-JUN-1999;   99US-0139152P,
PR  18-JUN-1999;   99US-0139453P,
PR  18-JUN-1999;   99US-0139492P,
PR  18-JUN-1999;   99US-0139454P,
PR  18-JUN-1999;   99US-0139455P,
PR  18-JUN-1999;   99US-0139456P,
PR  18-JUN-1999;   99US-0139457P,
PR  18-JUN-1999;   99US-0139458P,
PR  18-JUN-1999;   99US-0139459P,
PR  18-JUN-1999;   99US-0139460P,
PR  18-JUN-1999;   99US-0139461P,
PR  18-JUN-1999;   99US-0139462P,
PR  18-JUN-1999;   99US-0139463P,
PR  21-JUN-1999;   99US-0139750P,
PR  22-JUN-1999;   99US-0139763P,
PR  23-JUN-1999;   99US-0139817P,
PR  23-JUN-1999;   99US-0139899P,
PR  24-JUN-1999;   99US-0140353P,
PR  28-JUN-1999;   99US-0140354P,
PR  29-JUN-1999;   99US-0140695P,
PR  30-JUN-1999;   99US-0140823P,
PR  01-JUL-1999;   99US-0140991P,
PR  01-JUL-1999;   99US-0141287P,
PR  02-JUL-1999;   99US-0141842P,
PR  06-JUL-1999;   99US-0142055P,
PR  08-JUL-1999;   99US-0142154P,
PR  09-JUL-1999;   99US-0142390P,
PR  12-JUL-1999;   99US-0142803P,
PR  13-JUL-1999;   99US-0142920P,
PR  14-JUL-1999;   99US-0142977P,
PR  15-JUL-1999;   99US-0143542P,
PR  16-JUL-1999;   99US-0143624P,
PR  16-JUL-1999;   99US-0144005P,
PR  19-JUL-1999;   99US-0144085P,
PR  19-JUL-1999;   99US-0144086P,
PR  19-JUL-1999;   99US-0144325P,
PR  19-JUL-1999;   99US-0144331P,
PR  19-JUL-1999;   99US-0144332P,
PR  19-JUL-1999;   99US-0144333P,
PR  19-JUL-1999;   99US-0144334P,
PR  20-JUL-1999;   99US-0144335P,
PR  20-JUL-1999;   99US-0144352P,
PR  20-JUL-1999;   99US-0144632P,

PR  20-JUL-1999;   99US-0144884P,
PR  21-JUL-1999;   99US-0144814P,
PR  21-JUL-1999;   99US-0145086P,
PR  21-JUL-1999;   99US-0145088P,
PR  22-JUL-1999;   99US-0145089P,
PR  22-JUL-1999;   99US-0145085P,
PR  22-JUL-1999;   99US-0145087P,
PR  22-JUL-1999;   99US-0145089P,
PR  22-JUL-1999;   99US-0145091P,
PR  23-JUL-1999;   99US-0145145P,
PR  23-JUL-1999;   99US-0145218P,
PR  23-JUL-1999;   99US-0145224P,
PR  26-JUL-1999;   99US-0145276P,
PR  27-JUL-1999;   99US-0145913P,
PR  27-JUL-1999;   99US-0145918P,
PR  28-JUL-1999;   99US-0145919P,
PR  02-AUG-1999;   99US-0145951P,
PR  02-AUG-1999;   99US-0146386P,
PR  02-AUG-1999;   99US-0146388P,
PR  02-AUG-1999;   99US-0146389P,
PR  03-AUG-1999;   99US-0147038P,
PR  04-AUG-1999;   99US-0147202P,
PR  04-AUG-1999;   99US-0147303P,
PR  05-AUG-1999;   99US-0147192P,
PR  05-AUG-1999;   99US-0147260P,
PR  06-AUG-1999;   99US-0147493P,
PR  06-AUG-1999;   99US-0147303P,
PR  09-AUG-1999;   99US-0147416P,
PR  09-AUG-1999;   99US-0147493P,
PR  10-AUG-1999;   99US-0147935P,
PR  11-AUG-1999;   99US-0148171P,
PR  12-AUG-1999;   99US-0148319P,
PR  13-AUG-1999;   99US-0148341P,
PR  13-AUG-1999;   99US-0148565P,
PR  16-AUG-1999;   99US-0148684P,
PR  17-AUG-1999;   99US-0149368P,
PR  18-AUG-1999;   99US-0149175P,
PR  18-AUG-1999;   99US-0149426P,
PR  20-AUG-1999;   99US-0149722P,
PR  20-AUG-1999;   99US-0149723P,
PR  20-AUG-1999;   99US-0149929P,
PR  23-AUG-1999;   99US-0149930P,
PR  25-AUG-1999;   99US-0149902P,
PR  26-AUG-1999;   99US-0150084P,
PR  27-AUG-1999;   99US-0150566P,
PR  27-AUG-1999;   99US-0151065P,
PR  30-AUG-1999;   99US-0151066P,
PR  31-AUG-1999;   99US-0151080P,
PR  01-SEP-1999;   99US-0151303P,
PR  07-SEP-1999;   99US-0151438P,
PR  10-SEP-1999;   99US-0151930P,
PR  13-SEP-1999;   99US-0152363P,
PR  15-SEP-1999;   99US-0153070P,
PR  16-SEP-1999;   99US-0153758P,
PR  20-SEP-1999;   99US-0154018P,
PR  22-SEP-1999;   99US-0154039P,
PR  23-SEP-1999;   99US-0154779P,
PR  24-SEP-1999;   99US-0155139P,
PR  28-SEP-1999;   99US-0155486P,
PR  29-SEP-1999;   99US-0155559P,
PR  04-OCT-1999;   99US-0156458P,
PR  05-OCT-1999;   99US-0155996P,
PR  06-OCT-1999;   99US-0157117P,
PR  06-OCT-1999;   99US-0157753P,
PR  07-OCT-1999;   99US-0157865P,
PR  08-OCT-1999;   99US-0158029P,
PR  12-OCT-1999;   99US-0158232P,
PR  12-OCT-1999;   99US-0158369P,
PR  13-OCT-1999;   99US-0159293P,
PR  13-OCT-1999;   99US-0192293P,
PR  13-OCT-1999;   99US-0192294P,
PR  14-OCT-1999;   99US-0192295P,
PR  14-OCT-1999;   99US-0159229P,
PR  14-OCT-1999;   99US-0159330P,
PR  14-OCT-1999;   99US-0159331P,
PR  14-OCT-1999;   99US-0159637P,
```

```
PR   14-OCT-1999;   99US-0159638P.
PR   18-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.
```

```
Query Match        19.8%;    Score 141.5;  DB 3;   Length 177;
Best Local Similarity  28.8%;   Pred. No. 1.9e-06;
Matches   32;   Conservative  30;  Mismatches  44;   Indels   5;   Gaps   1;

Qy    35  MMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELRIRVYM   94
              |::::|  |:|  :| |:||:||       ::: :|   ||:| :|  ||||||||||
Db     1  MLTLAEPEVLAAMQPLQRAKTMHLLAEATTLYELRLRCTGVDPDDHRVKSEIERINVYR   60

Qy    95  NRVKEITDKKK------AGKLDRGAASRFVKNALWEPKSKNASKVANKGKSK   140
              : ::           :| |:|||:|:|:::|||:| : :| :::
Db    61  EKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSLPDLTSTQKQSIRDLSKGE   111
```

Search completed: January 4, 2005, 13:42:24
Job time : 78 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 4, 2005, 07:11:31

Title: US-09-701-618A-1

Perfect score: 1156

Sequence: 1 cttcccgggactggagtc.........aggctgaaatgttgtaaaa 1156

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Search time 4255.51 Seconds
(without alignments)
9898.769 Million cell updates/sec

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

	Result No.	Score	Query Match Length	DB	ID	Description
	1	765.2	988	4	BM923929	BM923929 AGENCOURT
	2	650.6	915	4	CN645648	CN645648 ILLUMIGEN
	3	616	993	7	CO647214	CO647214 ILLUMIGEN
	4	606.4	880	4	BG193528	BG193528 RST12662
	5	596.4	1014	4	BM451300	BM451300 AGENCOURT
	6	592.2	847	4	BG189347	BG189347 RST8391 A
	7	590.2	853	4	BG217959	BG217959 RST37683
	8	584.6	849	4	BG185663	BG185663 RST4614 A
	9	584.6	878	5	BI753996	BI753996 60302743I
	10	579.6	781	5	BU854138	BU854138 AGENCOURT
	11	575.4	887	4	AL540005	AL540005 AGENCOURT
	12	568	807	4	BM679939	BM679939 UI-E-EO1-
	13	567	909	4	AI800532	AI800532 tc12a11.x
	14	564.8	1209	4	BG213559	BG213559 RST33166
	15	562.4	859	4	BG206383	BG206383 RST25831
	16	551.6	782	5	BU852693	BU852693 AGENCOURT
	17	543.4	879	4	BG210469	BG210469 RST30094
	18	534.4	642	4	BM973398	BM973398 UI-CF-EC1
	19	532.8	575	1	AI553826	AI553826 tn29608.x
	20	532.8	579	1	AI004680	AI004680 ou59d03.x
	21	532.8	609	4	CA943561	CA943561 ir29g10.x
	22	532.8	640	5	BU675203	BU675203 UI-CF-DU1
	23	532.8	653	5	BU628390	BU628390 UI-H-FL0-
	24	532.8	713	2	AW960444	AW960444 EST372515
c	25	531.8	609	1	AI809071	AI809071 wf68a10.x
c	26	531.8	806	4	BI836412	BI836412 603082950
c	27	531.2	571	1	AI379715	AI379715 tc41a10.x
c	28	531.2	607	1	AI800846	AI800846 wg13h12.x
c	29	524.4	642	6	CA314481	CA314481 UI-CF-FN0
c	30	524.2	716	7	CN342947	CN342947 17000s999
c	31	522.2	544	2	AW024957	AW024957 wu70c05.x
c	32	522	788	1	AV763046	AV763046 AV763046
c	33	521.4	571	1	AI123353	AI123353 qa48d09.x
c	34	521.4	535	1	AI435863	AI435863 th79h07.x
c	35	521.4	831	4	BG185119	BG185119 RST4077 A
c	36	520.8	539	2	BE857571	BE857571 7g02h10.x
c	37	520	548	2	AW118590	AW118590 xd94c09.x
c	38	515.4	586	4	BM675402	BM675402 UI-E-EJ0-
c	39	515.2	813	5	BU568056	BU568056 AGENCOURT
c	40	510.6	654	1	AV686813	AV686813 AV686813
c	41	510.6	738	5	BU561215	BU561215 AGENCOURT
c	42	508.4	535	2	BF434793	BF434793 7o72f11.x
c	43	506.2	963	6	CD515525	CD515525 AGENCOURT
c	44	499.2	838	6	BG210470	BG210470 RST30095
c	45	494.6	602	1	AV685739	AV685739 AV685739

---

us-09-701-618a-1.rst

ALIGNMENTS

RESULT 1
BM923929
LOCUS   BM923929   988 bp   mRNA   linear   EST 12-MAR-2002
DEFINITION AGENCOURT_6709679 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5760072
            5', mRNA sequence.
ACCESSION   BM923929
VERSION   BM923929.1 GI:19374308
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 988)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12806 row: i column: 01
            High quality sequence stop: 722.
FEATURES   Location/Qualifiers
     source   1..988
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5760072"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female. Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon cloning). Average insert size 1.4 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber (Invitrogen). Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
ORIGIN

RESULT 2

```
LOCUS       CN645648                915 bp    mRNA    linear   EST 13-MAY-2004
DEFINITION  ILLUMIGEN_MCO_24688 Katze_MMPL2 Macaca mulatta cDNA clone
            IBIUW:10464_57 similar to Bases 9 to 906 highly similar to human
            CID (Hs.380759), mRNA sequence.
ACCESSION   CN645648
VERSION     CN645648.1  GI:47159091
KEYWORDS    EST.
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
  TITLE     Large-scale Rhesus Macaque cDNA Sequencing
  JOURNAL   Unpublished (2003)
COMMENT     Contact: C. Magness
            Illumigen Biosciences Inc.
            2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
            Tel: 2063780400
            Fax: 2063780408
            Email: cmagness@illumigen.com
            Sequenced on 2004.03.02. 821 020 bases.
            PCR Primers
            FORWARD: CCCTCACTAAAGGGAACAAAA
            BACKWARD: CACTATAGGGCGAATTGGGTA
            Insert Length: 915   Std Error: 0.00
            Plate: CL000128 row: A  column: 07
            Seq primer: CCCTCACTAAAGGGAACAAAA
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..915
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /strain="Indian"
                     /db_xref="taxon:9544"
                     /clone="IBIUW:10464"
                     /sex="male"
                     /dev_stage="newborn infant"
                     /clone_lib="Katze_MMPL2"
                     /lab_host="E. coli SOLR"
                     /note="Organ: placenta; Vector: Uni-ZAP XR; Site_1: EcoR
                     I; Site_2: Xho I; Created from Stratagene ZAP-cDNA
                     Synthesis kit (catalog #200400) and ZAP-cDNA Gigapack III
                     Gold Cloning Kit (Catalog #200450)"
ORIGIN
```

Query Match   56.3%;   Score 650.6;   DB 7;   Length 915;
Best Local Similarity   86.9%;   Pred. No. 3.6e-126;
Matches 792;   Conservative   0;   Mismatches 44;   Indels 75;   Gaps   4;

```
Qy   100  GGAGCAGAGCTGGGCCATAATGGGCAGGTGAAGAAAATTAATGAAGACTATCCAGTAGAAATT  159
          |||||||||  ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     5  GGCCCTGAGTCAGCCCATAATGGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATT   64

Qy   160  CACGAGTATTTGTCAGCCGTTTGAGAATTCCATTGGTCGTCGTGGATGAGATCGCTGAAGACC  219
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    65  CATCAGTATTTGTCAGCCGTTTGAGAATTCCATTGGTCGTCGTGGATGAGATCGCTGAAGACC  124

Qy   220  ATGATGTCTGTTTCTAGAAAATGAGTTGTTGCAGAAGTAGTTGGACAACAGCAAAA  279
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   125  ATGATGTCTGTTTTCTAGAAAATGAAGTGTTGCAGAAGTAGTTGGACCACAGCAAAA  184

Qy   280  GTGGATTTGGTTTTCTCGCATACACATTAAATTCAAATGTTTTTGGGTTTATTTG  339
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   185  GTGGATTTGGTTTTCTCCGCATACACATTAAATTCAAATGTTTTTGGGTTTATTTG  244

Qy   340  GGAGTTAATCCTTAAGGAAACATCCAGTAAAACAGGAATTCAGAGTTATATATG  399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245  GGAGTTAATCCTTAAGGAAACATCCAGTAAAACAGGAATTCAGAGTTATATATG  304
```

Query Match   66.2%;   Score 765.2;   DB 4;   Length 988;
Best Local Similarity   90.6%;   Pred. No. 3.1e-150;
Matches 885;   Conservative   0;   Mismatches 26;   Indels 66;   Gaps   4;

```
Qy    91  GGAGGGGTGAGGAGCAGAGCTGGGCCATAATGGGCAGGTGAGGAAGAAATTAATGAAGACTATCCA  150
          ||||          ||||||||||||||||||||||||   |||||||||||||||||||||||
Db     8  GGAGTCGGAAGGCCCGTGAGTCAGCCCATAATGGGCAGGTGAAGAAATTAATGAAGACTATCCA   67

Qy   151  GTAGAAATTCACGAGTATTTGTCAGCCGTTTGAGAATTCCATTGGTCGTCGTGGATGATG  210
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    68  GTAGAAATTCACGAGTATTTGTCAGCCGTTTGAGAATTCCATTGGTCGTCGTGGATGAGATG  127

Qy   211  CTGAAGACCATGATGTCTGTTTCTAGAAAATGAGTTGTTGCAGAAGTTGGATCCACTTGAA  270
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128  CTGAAGACCATGATGTCTGTTTCTAGAAAATGAGTTGTTGCAGAAGTTGGATCCACTTGAA  187

Qy   271  CAAGCAAAAGTTGGATTTGGTTTTCTCGCATACACATTAAATTCAAATGTTTTTGGGTTTATTTG  330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   188  CAAGCAAAAGTGGATTTGGTTTTCTCGCATACACATTAAATTCAAATGTTTTTGGGTTTATTTG  247

Qy   331  GCAACCCAAGGAGTTAATCCTTAAGGAAACATCCAGTAAAACAGGAATTCAGAGTTATATATG  390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248  GCAACCCAAGGAGTTAATCCTTAAGGAAACATCCAGTAAAACAGGAATTCAGAGTTATATATG  307

Qy   391  GTATATATGAACAGAGTCAAGGAAATAACAGAGAAAAGGCTGGGCAAGCTGGGACAGA  450
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   308  GTATATATGAACAGAGTCAAGGAAATAACAGAGAAAAGGCTGGGCAAGCTGGGACAGA  367

Qy   451  GGTGCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGAACCAAAATCGAAAAAATGCATCA  510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   368  GGTGCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGAACCAAAATCGAAAAAATGCATCA  427

Qy   511  AAAGTTGCCAATAAAGGAAGAAAAGTAAACAGGAATTCAACTTTTTGGGTTTTGATTTATTTG  570
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   428  AAAGTTGCCAATAAAGGAAGAAAAGTAAACAGGAATTCAACTTTTTGGGTTTTGATTTATTTC  487

Qy   571  AAAAAGTAC-----------------------------------  579
          |||||||||
Db   488  AAAAAGTACATCTTCCCCCACCCCCCGCAAAATAATTCTGTGGCAGGGACAGGGTTAAA  547

Qy   580  ------------ATTAATATGTAAATCACA---GTAATATGTAAAGCTAAAATAC--TTCCTCTC  625
                      |||||||||||||| ||
Db   548  TGTGTTCTTTATTTAATATGTTAAATTCACAGTAATATGTAAAGCTAAAATACTTCCTCTC  607

Qy   626  CAAAGACATTATTCTTTAATTACACTGAGGATTTTAACATTGTGATATATTATATA  685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   608  CAAAGACATTATTCTTTAATTACACTGAGGATTTTAACATTGTGATATATTATATA  667

Qy   686  TTTATAAATTTTACCCATCTCTTGATGAGACTCTTATTTCTTTATAATAGGNCAGTCTTGCAAG  745
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   668  TTTATAATTTTACCCATCTCTTGATGAGACTCTTATTTCTTTATAATAGGNCAGTCTTGCAAG  727

Qy   746  TACCATTTTATAAGCAGCTGTGAAATTTAAGTGAAATGTTCTTTGTGTAAACATTGTACTA  805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   728  TACCATTTTATAAGCAGCTGTGAAATTTAAGTGAAATGTTCTTTGTGTAAACATTGTACTA  787

Qy   806  TTTTAAATGAATAAATGACCTTATCGTATCGTGATGCTGAAATATAGGTACAT  865
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   788  TTTTAAATGAATAAATGACCTTATCGTATCGTGATGCTGAAATATAGGTACAT  847

Qy   866  CTGTTTTCACTATATGATATTAAGAACAACGGTG-AATGACTTAAAATGTTCAATTTTTTTCTG  924
          ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db   848  CTGTTTTCACTATATGATATTAAGAAACGGTGAAATGACTTAAAATGTTCAATTTTTTTCTG  907

Qy   925  TATAGATACTTTTATCATGTTTCATGAGTTTTAGGAAATACTGCTTTAAAAATGTCATTTTTTCTG  984
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   908  GATAGATACTTTTATCATGTTTTCATGAGTTTTAGGAATTACTGCTTTGNTGATATTTCAATG  967

Qy   985  GATAGTAAACTAAAAGTTT 1001
          |||||||||||||||||
Db   968  TGTGAAACTAAAAACTT  984
```

```
RESULT 3
CO647214
LOCUS       CO647214              993 bp    mRNA      linear   EST 23-JUL-2004
DEFINITION  ILLUMIGEN_MCQ_40181 Katze MMPB2 Macaca mulatta cDNA clone
            IBIUW:23958 5' similar to Bases 359 to 913 highly similar to human
            CID (Hs.380759), mRNA sequence.
ACCESSION   CO647214
VERSION     CO647214.1  GI:50568708
KEYWORDS    EST.
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
  TITLE     Large-scale Rhesus Macaque cDNA Sequencing
  JOURNAL   Unpublished (2003)
COMMENT     Contact: C. Magness
            Illumigen Biosciences Inc.
            2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
            Tel: 206780400
            Fax: 206780408
            Email: cmagness@illumigen.com
            Sequenced on 2004.05.28. 778 Q20 bases. Library Preparation: Prof.
            Michael Katze Lab at University of Washington DNA Sequencing;
            Illumigen Biosciences Inc. For further information, see
            http://www.macaque.org
                     PCR Primers
                     FORWARD: CCCTCACTAAAGGGAACAAA
                     BACKWARD: CACTATAGGGCGAATTGGGTA
                     Plate: CL000339  row: G  column: 08
                     Insert Length: 993    Std Error: 0.00
                     Seq primer: CCCTCACTAAAGGGAACAAA
                     POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..993
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /strain="Indian"
                     /db_xref="taxon:9544"
                     /clone="IBIUW:23958"
                     /sex="male"
                     /tissue_type="blood"
                     /cell_type="PBMC"
                     /dev_stage="adult"
                     /lab_host="Electromax DH10B"
                     /note="Vector: pDONR 222; Site 1: BsrG I; Site 2: BsrG I;
                     clone_lib="Katze MMPB2"
                     /note="Vector: pDONR 222; Site 1: BsrG I;
                     created from CloneMiner cDNA Library Construction kit
                     (catalog #18249-029)"
ORIGIN
Query Match     53.3%;  Score 616;  DB 7;  Length 993;
Best Local Similarity   85.5%;   Pred. No. 6.7e-119;
Matches  776;  Conservative   0;  Mismatches   55;  Indels  77;  Gaps   5;
```

```
Qy    400 AACAGAGTCAAGGAAATAACAGAACAAGAAAAAGGGCTGGGCAAGCTGGACAGCAGAGGTGCAGCT 459
          ||||||||||||||||||||||||
Db    305 AATAAAGGGAAAAAGTAAAAGTAACTAAAAAGGGCTGGGCAAGCTGGACAGCAGAGGTAC       364

Qy    460 TCAAGATTTGTAAAAAATGCCCTCTGGGACCAAAAAATCGAAAAAATCGATCACAGAAAAGTTGCC 519
          |||||||||||||||||||||||||||||||||||||||||
Db    365 TCAAGATTTGTAAAAAATGCCCTCTGGGACCAAAAAATCGAAAAAATGCATCAAAGAGTTGCC   424

Qy    520 AATAAAAGGAAAAAGTAAAAGTAACTAAAAAAGGGCTGGGCAAGCTGGACAGCAGAGGTAC     579
          ||||||||||||||||||
Db    425 AATAAAAGGAAAAAGTAAAAGTAACTAAAAAGGGCTGGGCAAGCTGGACAGCAGAGGTAC      484

Qy    580 ATTAATATGTAATCAC------------------------------------------------- 595
          ||||||||||||||||
Db    485 ATCTTTTTTTTTTCCCCTCCACAAAAATAATTAATTTTGGTTTTGATGTACACATATTCAAAAAGTAC 544

Qy    596 ------------------AGTAATATGTAAAGCTAAATAC-TTCCTCTCCA               627
                            |||||||||||||
Db    545 CTTATTGAATTAATATGTAAATTTACAGTAAATATGTAAAAGCTAAATACTATTTCCTTTCA     604

Db    605 AAGATGTGTATCTTTATTGATTTGCACTGAGGATTTTAACATTGTGATAATATATTTATTATTTT    664

Qy    628 AAGATCATTATATCTTTATTGAATTAACACTGAGGAGATTTAACATTGTGAATTTAAGAGTAATT  687
          |||||||||||||
Db    605 AAGATGTGTATCTTTATTGATTTGCACTGAGGATTTTAACATTGTGATAATATATTTATTATTTT    664

Qy    665 TATAATTTACCATCTCTTTTATATGAGAGACTCTTAATTCTTTATATAGGTCAATCTGCAAA     724
          |||||||||||||||
Db    665 AGTTGGATCCACTTGAACACAGCAGCAAAAGTGGAATTGGTTTCTGCCATACACATTAAATTCAA   313

Qy    688 TATAATTTACCATCTCTT--GATGAGACTCTTATTTCTTTATATAGGTCAGTCTGCAAG       745
          ||||||||||||||||||
Db    665 ...

Qy    725 TACCATTTTATAAGCAGCTGTGAAATTTAAGTAGAATAGTGTCTTGTTGAAACATTTGTACTA    784
          ||||||||
Db    746 TACCATTTTATAAGCAGCTGTGAAATTTAAGTAGAATAGTGTCTTGTTGAAACATTTGTACTA    805

Qy    806 TTTTAAAATGAATAATGAATGACCTTATGAAGTAGCTATTCTGTAGGCTGAAATATAAGTAGTACAT 865
          ||||||||||||||
Db    785 TTTTAAAATGAATAATGAATGACCTTATGATGTACTGTCTGTAGGCCAAAAATTATAAGTTTACAT   844

Qy    866 CTGTTTTCACTATATGATATTAAGAAAGGGTG-AATGACTTAAAATGTTCATTTTTTTTCTG      924
          |||||||||||||
Db    845 CTGTTTTCACTATATGACATTTAGAAAGACTTAAGGTTTAATGACTTAAATGTTCTTTTTTTTTCTG 904

Qy    925 TATAGATACTT  935
          ||||
Db    905 TAGTTATTTTT  915
```

```
Qy    254 AGTTGGATCCACTTGAACACAGCAGCAAAAGTGGAATTGGTTTCTGCCATACACATTAAATTCAA 313
          ||||||| ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     21 AGTTGGACCCACTTGAACAACAGCAAAAGTGGAATTGGTTTCTGCCATACACATTAAATTCAA    80

Qy    314 TGTTTTGGGTTTTATTGGGCAACCCAAGGAGTTAATCCTAAGGAGAACATCCAGTAAAACAGG   373
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     81 TGTTTTGGGTTTTATTGGGCAACTCAAGGAGTTAATCCTAAGGAGAACATCCAGTAAAACAGG   140

Qy    374 AATTGAAAGAATCAGATGTTATATATATGAACACAGAGTCAAGGAGAAATAACAGACGAGAAGG  433
          ||||||||||||||||     |||||||||||||||||||||||||||||||||||||||
Db    141 AATTGAAAGAATCAGATATATTATATATGAACACAGAGTCAAGGAGAAATAACAGACGAGAAGG   200

Qy    434 CTGGCCAAGCTGGACAAGGTTGCAGCTTCAAGATTTGTAAAAATGCCCTCTGGGGAACCAA     493
          |||||||||||||| |||||||||||||||
Db    201 CTGGCAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGGAACCAA     260

Qy    494 AATCGAAAAAATGCATCAAAGAGTTGCCAATAAAAGGAAAAAAGTTAACTTTTTTGGTT     553
          |||||||||||||||
Db    261 AACCAAGAAAATGCATCAAAAGTTGCCAATAAAGGTAAAAGTAAAAGTTAACTTTTTTTGGTT    320

Qy    554 TTGATGTACACATATTCAAAAAAGTACATTAATAAAAAGTACATTAAATCTTTTATTTCAC---- 595
          |||||||||||||||
Db    321 TTGATGTACACATATTCAAAAAAGTACATTAATAAAAAGTACATTAAAT                 380

Qy    596 ------------------------------------------------AGTAATA          602
                                                            ||||||
Db    381 TGGGCAGGACAAGGTTAAATGTATTTTCTTATTGAATTAATATGTAAATTTACAGTAAATA     440

Qy    603 TGTAAAGCTAAATAC-TTCCTCTCCAAAAGATCATTATATCTTTATTGAATTAACACTGAGGAT  661
          |||||||||||||||
Db    441 TGTAAAGCTAAATAC TTCCTCTCCAAAAGATCATTATTATCTTTATTTAGTTACACTGAGGAT   500

Qy    662 TTTAACATTGTGAATATAATTTTATATTTTATTACCATCATCTC--TTGATGAGACTCTTAT  719
          |||||||||||||
Db    501 TTTAACATTGTGATATATATATAATTTTACCATCATCTCTTTATATTTTATTACCATCATCTTAT  560

Qy    720 TTCTTTATATAGGTCAGTCTGCAAGTACCATTTTATAAGCAGCTGTGAAATTTAAGTGA      779
          ||||
Db    561 TTCTTTATAGGTCAATCTGCAAATACCATTTTATAAGCAGCTGTGAAATTTAAGTGA       620

Qy    780 AATGTTCTTTGTAAACATTTGTACTATTTTAAAATGAATAATGAACTTATGAAGTATGCTA    839
          |||
Db    621 AATGTTCTTTGTAAACATTTGTACTATTTTAAAATGAATAATGACCTTATGATGTATGCTG    680
```

RESULT 4

```
BG193528
LOCUS       BG193528                880 bp    mRNA    linear    EST 21-APR-2001
DEFINITION  RST12662 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG193528
VERSION     BG193528.1  GI:13715215
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 880)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE   21227151
  PUBMED    11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 482.
FEATURES             Location/Qualifiers
     source          1..880
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="HT1080"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
ORIGIN
```

Qy 840 TCTGTAGGCTGAAATTATAGGTACACATCGTGTTTCACTAATAGTGATGATTAATTAAGAAAAAGGGTGA- 898
         |||||||||||||||| |||||||| ||||||||||||||||||||||||||||||||||||||||| |||
Db 681 TCTGTAGGCCAAAATTATAGGTACACTCGTGTTTCACTATATGATGAATTAAGAAAGCAATGAG 740

Qy 899 ATGACTTAAAAGTTCATTTTTTTTTCTGTATAGATACTTATCATGTTGTTATGATTTAGG 958
         ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db 741 ATGACTTAAAAGTTCATTTTTTTCTG---TAGTTATTTATCATGAACATGATTTAGG 798

Qy 959 AATTACTGCTTTGTTGGATATTCAAAAGTGTGAAAACTTAAAAGTTATGGGTTTGCTTACTTTTAATT 1018
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 799 AATTACTGCTTTGTTGGATATTCAAGCTGTGAAAACTTAAAAGTTATGGGTTTGCTTACTTTTAATT 858

Qy 1019 CTTGGCATGTTGCCCTCTATGTCCCATTTAAAATAAAAATACAATTCTCATTAACTTTAGATG 1078
         |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db 859 CTTGGCATGTTGCCCTCTATGTCCCATTTAAAATAAAAATACAATTCTCATTAACTTTAAAA 918

Qy 1079 GGAAATAA 1086
         |||| ||
Db 919 AAAAAAAA 926

RESULT 5

```
BM451300
LOCUS       BM451300               1014 bp    mRNA    linear    EST 05-FEB-2002
DEFINITION  AGENCOURT_6386679 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493810
            5', mRNA sequence.
ACCESSION   BM451300
VERSION     BM451300.1  GI:18500340
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1014)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE
```

Qy 316 TTTTGGGTTTATTTGGGCAACCCAAGGAGGTTAATCCTAAGGAACATCCAGTAAAACAGGGA 375
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 64   TTTTGGGTTTATTTGGGCAACCCAAGGAGGTTAATCCTAAGGAACATCCAGTAAAACAGGAA 123

Qy 376 TTGGAAAGAATCAGAGTATATGAACAGAGTGTCAAGAGTTCAAGGAGAAAATAACAGCAAGAAAAAGGCT 435
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 TTGGAAAGAATCAGAGTATATGAACAGAGTCAAGGAGAAAATAACAGCAAGAAAAAGGCT 183

Qy 436 GGCAAGCTGGACAGGAGGTGCAGCTTCAAGATTGTAAAAAATGCCCTCTGGGAACCAAAA 495
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 GGCAAGCTGGACAGGAGGTGCAGCTTCAAGATTGTAAAAAATGCCCTCTGGGAACCAAAA 243

Qy 496 TCGGAAAAATGCATCAAAAGTTCCAATAAAAGGAAAAAAGTAAAAGTTAACTTTTTTTGTTTT 555
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 TCGGAAAAATGCATCAAAAGTTCCAATAAAAGGAAAAAAGTAAAAGTTAACTTTTTTGTTTT 303

Qy 556 GATGTACACATATTCAAAAAAGTAC---------------------------------- 579
         |||||||||||||||||||||||||
Db 304 GATGTACACATATTCAAAAAAGTACATCTTCCCCCCCCCGCCGGCAAAATAAATTCTGTGG 363

Qy 580 ----------------------ATTAATATGTAATCACA---GTAATATGTAAAGC 610
                               ||||||||||||||||||| |||||||||||||
Db 364 CAGGGCAAGGTTTAAATGTGTTCTTATTAATATGTAATCACAGTAATATGTAAAGC 423

Qy 611 TAAAATAC-TTCCTCTCCAAAGATACAATTACTCTTTATTAATTAGCACTGAGGATTTTAACAT 669
         ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 424 TAAAATACTTTCCTCTCCAAAGATACAATTACTCTTTATTGATTAGCACTGAGGATTTTAACAT 483

Qy 670 TGTGATATATATAATATTTAATTACCCATCTCTTTGATGAGAGACTCTTATTTCTTTATAT 729
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 484 TGTGATATATATAATATTTAATTACCCATCTCTTTGATGAGAGACTCTTATTTCTTTATAT 543

Qy 730 AGGTCAGTCTTGCAAGTACCCATTTTATAAGCAGCGTGTGGAAAATTAAGTGAAAATGTTCTTT 789
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 544 AGGTCAGTCTTGCAAGTACCATTTNTATAAGCAGCTGTGGAAAATTAAGTGAAAATGTTCTTT 603

Qy 790 GTAAACAATTGTACTATTTAAATGAAATAAATGACCTTAATGAAGTATGCTATCTGTAGGCT 849
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 604 GTAAACAATTGTACTATTTAAATGAAATAAATGACCTTAATGAAGTATGCTATCTGTAGGCT 663

Qy 850 GAAATTATAGGTACATCTGTTTCACTATATGATATTAAGAAAGCGTGAAATGACTTAAAT 909
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 664 GAAATTATAGGTACATCTGTNTCACTATATGATATTAAGAAAGCGTGAAAATGACTTAA 723

Qy 910 GTTCATTTTTTTCTGTATAGATACTTATCATGTTTCATGATTTTTAGGGAATTAACTGCTT 969
         ||||||||||||||||||||| |||||||||||||||||||||||||||||| |||||||
Db 724 TGTCATTTTTTTCTGATAGATACTTTATCATGGTTTCATGAATTTTAGG-ATTACTGCTT 782

Qy 970 TGTTGGATATTCAAAAGTGTGAAACTAAAGTTTATGGTTGTACTTTTAATTCTTGGCCATGTT 1029
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 783 TGTTGGATATTCAAAAGTGTGAAACTAAAACTTTATGGGTTGGACTTT-ATTCTTTGCATGTT 841

Qy 1030 GCCCTCTATGTCCCATTTAAAATAAAAATACAATTCCATTAAT 1069
         ||||||||||||||||||||||||||||||||||||||||||||
Db 842 G-CTCTAGTCCATTAAAATAAAAATACATCCTATTAACTTA 880

ORIGIN

Qy 256 TTGGATCCCTTGAACAAGCAAAATAGTGGATTGGTTTCTGCATACACATTAAATTCAAATG 315
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4   TTGGATCCCTTGAACAAGCAAAATAGTGGATTGGTTTCTGCATACACATTAAATTCAAATG 63

Qy Query Match   52.5%;   Score 606.4; DB 4;   Length 880;
    Best Local Similarity   86.7%;   Pred. No. 6.9e-117;
    Matches 763; Conservative   0; Mismatches   48;   Indels 69;   Gaps   6;

JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LIAMI2118 row: b column: 19
          High quality sequence stop: 606.

FEATURES   Location/Qualifiers
source   1..1014
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="IMAGE:5493810"
           /tissue_type="retinoblastoma"
           /lab_host="DH10B (phage-resistant)"
           /clone_lib="NIH_MGC_67"
           /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
           Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
           Average insert size 1.75 kb. Library constructed by Life
           Technologies."

ORIGIN

Query Match   51.6%; Score 596.6; DB 4; Length 1014;
Best Local Similarity   89.2%; Pred. No. 7.9e-115;
Matches 725; Conservative 0; Mismatches 21; Indels 67; Gaps 5;

```
Qy   91 GGAGGGTGAGGAGGAGGAGCTGGCCATAAATGCAGGTGAGAGAAATTAATAGAACGTATCCA 150
        |||                 |||||||||||||||||||||||||||||||||||||||||||
Db   33 GGAGTCGGAAGGCCGTGAGTCAGCCCATAATGGCTACGGTGAAGAAATTAATTAATGAAGACTATCCA 92

Qy  151 GTAGAAATTCACGAGTATTTGTCACCGTTTGAGGAATTCCATTGGTGTCGTGTGGATGATGAGATG 210
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   93 GTAGAAATTCACGAGTATTTGTCACCGTTTGAGGAATTCCATTGGTGTCGTGTGGATGATGAGATG 152

Qy  211 CTGAAGACCATGATGTCTCGTTTCTTAGAAATTGAGTTGTTGCGAGAAGTTGGATCCACTTGAA 270
        |||||||||||||||||||||||||||||||||||||||||||||          |||||||||
Db  153 CTGAAGACCATGATGTCTCGTTTCTTAGAAATTGAGTTGTTGGATCCACTTGAA 212

Qy  271 CAAGCAAAAGTGGATTGGTTTCTGCATACACATTAAATTCAATGTTTTGGGTTTATTTG 330
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  213 CAAGCAAAAGTGGATTGGTTTCTGCATACACATTAAATTCAATGTTTTGGGTTTATTTG 272

Qy  331 GCAACCCAAGGAGTTAATCCTAAGGAACAATCCAGTAAAAACAGGAATCAGA 390
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  273 GCAACCCAAGGAGTTAATCCTAAGGAACAATCCAGTAAAAACAGGAATTCAGA 332

Qy  391 GTAATAATGAACAGAGTCAAGGAAATAACAGACAAGAAAAGGCTGGCAAGCTGGACAGA 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  333 GTAATAATGAACAGAGTCAAGGAAATAACAGACAAGAAAAGGCTGGCAAGCTGGACAGA 392

Qy  451 GGTGCAGCTTCAAGATTTGTAAAAAAATGCCCTCTGGGAACCAAAATCGAAAAATGCATCA 510
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  393 GGTGCAGCTTCAAGATTTGTAAAAAAATGCCCTCTGGGAACCAAAATCGAAAAATGCATCA 452

Qy  511 AAAGTTGCCAATAAAGGAAAAAGTAAAAGTTAAACTTTTTGGTTTTGATGTACACATATTC 570
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  453 AAAGTTGCCAATAAAGGAAAAAGTAAAAGTTAAACTTTTTGGTTTTGATGTACACATATTC 512

Qy  571 AAAAAGTAC--------- 579
        |||||||||
Db  513 AAAAAGTACATCTTCCCCACCCCCCGCAAAAAAATAATTCTGTGCAGGCAAGGTTTANA 572

Qy  580 ----------ATTAATATGTAAATCACA---GTAATATGTAAAGTAAATAC-TTCCTCTC 625
                  |||||||||||||||||||   ||||||||||||||||||| ||||||||
Db  573 TGTGTTTCTTATTAATATGTAAATTCACAGTAAATATATGTAAATTACTTCCTCTC 632

Qy  626 CAAAGATCATTATCTTTATTGATTAGCACTGAGGATTTTAACACTTGTGATATATATATA 685
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  633 CAAAGATCATTATCTTTATTGATTAGCACTGAGGATTNTAACATTGNGATATATATATA 692
```

RESULT 6

COMMENT   Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 514.

FEATURES   Location/Qualifiers
source   1..847
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /cell_line="HT1080"
           /clone_lib="Athersys RAGE Library"
           /note="See 'Creation of Genome-wide Protein Expression
           Libraries using Random Activation of Gene Expression',
           Nature Biotechnology, in press. Note that even though the
           cell type indicated is HT1080, since a random activation
           method was used, these sequence tags are not necessarily
           expressed in HT1080 under normal circumstances."

ORIGIN

Query Match   51.2%; Score 592.2; DB 4; Length 847;
Best Local Similarity   89.2%; Pred. No. 6.6e-114;
Matches 734; Conservative 0; Mismatches 18; Indels 71; Gaps 6;

```
Qy  279 AGTGGATTGGTTTCTGCATACACATTAAATTCAATGTTTTGGGTTATTTGGCAACCCA 338
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 AGTGGATTGGTTTCTGCATACACATTAAATTCAATGTTTTGGGTTATTTGGCAACCCA 84

Qy  339 AGGAGTTAATCCTAAGGAACAATCCAGTAAAAACAGGAATTCAGAGTAATAATGAGTATAT 398
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 AGGAGTTAATCCTAAGGAACAATCCAGTAAAAACAGGAATTCAGAGTAATAATGAGTATAT 144

Qy  399 GAACAGAGTCAAGGAAATAACAGACAAGAAAAGGCTGGCAAGCTGGACAGAGGTATGCAGC 458
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  399 GAACAGAGTCAAGGAAATAACAGACAAGAAAAGGCTGGCAAGCTGGACAGAGGTATGCAGC 458
```

```
RESULT 7

LOCUS       BG217959      853 bp    mRNA    linear    EST 21-APR-2001
DEFINITION  RST37683 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG217959
VERSION     BG217959.1  GI:13743980
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
   MEDLINE  21227151
    PUBMED  11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 550.
FEATURES             Location/Qualifiers
     source          1..853
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."

ORIGIN

  Query Match       51.1%;   Score 590.2;   DB 4;   Length 853;
  Best Local Similarity   87.5%;   Pred. No. 1.7e-113;
  Matches 743;  Conservative   0;  Mismatches  33;   Indels  73;  Gaps  6;
```

```
Qy    145 GAACAGAGTCAAGGAAATAACACAGACAAGAAAAAGCTGCGCAAGCTGGACAGAGGTGCAGC 204
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    459 TTCAAGATTTGTAAAAAAATGGCCCTCTGGGA-ACCAAAAATCGAAAAAATGCATCAAAAGTTG 517

Qy    205 TTCAAGATTTGTAAAAAATCCCCTCTGGGATACCAAAAATCGAAAAAATGCATCAAAAGTTG 264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    518 CCAATAAAGGAAAAAGTAAAAAGTTAACTTTTTTCGTTTTGATTGTACACATATTCAAAAAGT 577

Qy    265 CCAATAAAGGAAAAAGTAAAAAGTTAACTTTTTTCGTTTTGATTGTACACATATTCAAAAAGT 324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    578 AC------------------------------------------------------ 579

Qy    325 ACATCTCCCCCCCCCCCCCCGCAAAATAATTCTGTGGCAGGGGCCAAGGTTTAAAATGTG 384
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 -----ATTAAATATGTAATCACA---GTAATATGTAAAGCTAAAATAC--TTCCTCCCAAA 629

Qy    385 TTTCTTATTAATATGTAAAATTCACAGTAAAATTAAAGCTAAAATACTTTCCTCTCCCAAA 444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    630 GATCATTATATCTTTTATTGATTAGCACTGAGGATTTTGTTGATTATTATATATATTTTA 689

Qy    445 GATCATTATATCTTTTATTGATTAGCACTGAGGATTTTGTTGATTATTATATATTTTA 504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    690 TAATTTACCACTCTCTTGATGAGACTCTTAATTCTTTATATAGGTCAGTCTTGCAAGTACC 749

Qy    505 TAATTTACCACTCTCTTGATGAGACTCTTAATTCTTTATATAGGTCAGTCTTGCAAGTACC 564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    750 ATTTTTATTAAGCAGCTGTGAAATTTTAAGTGAAAGTGTTCTTTGTAAACATTTGTACTATTTT 809

Qy    565 ATTTTTATTAAGCAGCTGTGAAATTTTAAGTGAAAGTGTTCTTTGTAAACATTTGTACTATTTT 624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    810 AAATGAATAATGACCTTATGGAAGTATGCTAATCTGTAGGCTGGAAATAATAACAAGTTGTACATCTGT 869

Qy    625 AAATGAATAATGACCTTATGGAAGTATGCTAATCTGTAGGCTGGAAATAATAACAAGTTGTACATCTGT 684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    870 TTTCACTATATGATATTAAGAAAGCGTG--AATGACTTAAAAGTTCATTTTTTTCTGTATA 928

Qy    685 TTTCACTATATGATATTAAGAAAGCGTGAAATGACTTAAAATGTTCATTTTTTTCTGATTA 744
          ||||||||||||||||||||||||  |||||||  ||||||||||||||||||||| ||
Db    929 GATACTTTATCATGTTTCATGATTTAAGGAATTACTGGCTTTTGTTGATATTCAAAAGTGTG 988

Qy    745 GATACTTTATCATGTTTCATGATTTAAGG-ATTACTGCTTGTTGTTGATATTCAAAAGTGTG 803
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    989 AAACTAAAAAGTTTATAGTTGTTGTACTTTTAATTCTTTATTCTCTGGCTGTGGC 1031

Qy    804 AAACTAAAAACTTTATGGGTGGGACCTTTATTCTATTCTCTGGCATTGTGTGGC 846
```

```
Qy    323 TTTTATTTGGGCAACCCAAGGAGAGTTAATCCTAAGGAACACATCCAGTTAAAACAGGAATTGGAAA 382
          |||| || |||| |||| ||||| ||||||||||||||||||||||  ||||||||| ||||
Db      4 TTTTATTTGCGCACCCAAGGAGTTAATCCTAAGGAACACATCCAGTTAAAACAGGAATTGGAAA 63

Qy    383 GAATCAGAGTATATATGAACAGAGTCAAGGAAATAACAGACAAGAAAAAGCTGGGCAAGC 442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 GAATCAGAGTATATATGAACAGAGTCAAGGAAATAACAGACAAGAAAAAGGCTGGGCAAGC 123

Qy    443 TGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAATGCCCTCTGGGGAACCAAAATCGAAAA 502
          ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db    124 TGGGCAGAGGTGCAGCTTCAAGATTTGTAAAAAAATGCCCTCTGGGGAACCAAAATCGAAAA 183

Qy    503 ATGGATCAAAAAGTTGCCAATAAAAGGAAAAAAGTAAAAAGTTAACTTTTTGGTTTGATGTAC 562
          |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db    184 ATGCATCAAAAAGTTGCCAATAAAGGAAAAAAGTAAAAAGTTAACTTTTGGTTTTGATGTAC 243

Qy    563 ACATATTCAAAAAGTAC-------------------------------------- 579
          ||||||||||||||||
Db    244 ACATATTCAAAAAGTACACATCTCCCCCCCCCCCCCCCCAAAAATAATTCTGTGTGCAGG 303

Qy    580 ------------------------------ATTAAATATGTAATCACA---GTAATATGTAAAGCTAA 614
                                        ||||||||||||||||||   |||||||||||||||
Db    304 GCAAGGTTTAAAATGTGTTTCTTATTAATATGTAATCACAGTAATATGTAAAGCTAA 363

Qy    615 TAC--TTCCTCTCCCAAAGATCATTATATCTTTTATTGATTAGCACTGAGGATTTTGTTG 673
          |||  ||  |||||||| ||||| ||||||||||||||||||||||||||||||||||
Db    364 TACTTTCCTCCCAAAGATCATTATATCTTTTATTGATTAGCACTGAGGATTTTGTTGTA 423

Qy    674 ATATTATATATATTTTTATTAAGTTCCATCTCCTTTGATGAGACTCTTAATTTCTTTATATAGGT 733
          |||||||||||||||  ||||||||||||| |||||||||||||||||||||||||||||
Db    424 ATATTATATATATTTAATTACCATCTCTTCCTTTGATGAGACTCTTAATTCTTTATATAGGT 483

Qy    734 CAGTCTTGCAAGTACCATTTTTATTAAGCAGCTGTGTGAAATTTAAGTGAAAGTTGTTCTTTGTAA 793
          |||||||||||||||||||||||||||||||||||||||  |||||||||||||| ||||
Db    484 CAGTCTTGCAAGTACCATTTTATTAAGCAGCTGTGAAATTTTAAGTGAAAGTTGTTCTTTGTAA 543

Qy    794 ACATTTGTACTATTTTAAATGAATAATGACCTTATGGAAGTATGCTAATCTGTAGGCTGGAAA 853
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 ACATTTGTACTATTTTAAATGAATAATGACCTTATGGAAGTATGCTAATCTGTAGGCTGGAAA 603

Qy    854 TTATAGGTACATCTGTTTTCACTATATGATATTAAGAAA-GCGTGAATGACTTAAAATGTT 912
          ||||||||||||||||||||||| |||||||||||||||  ||||| ||||||||||||||
Db    604 TTATAGGTACATCTGTTTTCACTATATGATATTAAGAAAGCGGGAAATGACTTAAAATGTT 663

Qy    913 CATTTTTTTCTGTATAGATACTTTATCATGTTTCATGATTTAAGGAATTA-CTGCTTTTG 971
          |||||||||||||||||||||||||||||||||||||||||||||||||   |||||||
Db    664 CATTTTTTTCTGTATAGATACTTTATCATGTTTCATGATTTAAGGAATTACTGCTTTTG 723

Qy    972 TTGATAATTCAAAAGTGTGAAAACTAAAAAGTTTATGGGTTGTGTAC--TTTAATTCTTGGCATGTT 1029
          |||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db    724 CTGAAAATTCAAAAGTGGGAAACTAAAAACTTTATGGGTTGACCTTTAATTACTGGCATGTT 783
```

```
Qy   1030 GCCCTATGTCCCATTTAAAATAAATAACATTCTCATTAACTTTAGATGGGAAATAAGGT 1089
Db    784 GCCCTATGTCCCATTTAAAATAAATAACTATTTTCCATTAACTCTAGATGGGAAATAAGT 843
          |||||    ||||||||||||||||||||    |||||||    |||||||||||||| ||

Qy   1090 TGTATGTTG 1098
Db    844 TTGATGTTG 852
          | ||||||

RESULT 8
BGI85663
LOCUS        BGI85663              849 bp    mRNA            EST 21-APR-2001
DEFINITION   RST4614 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION    BGI85663
VERSION      BGI85663.1  GI:13707350
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 849)
  AUTHORS    Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
             Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
             Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
             Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
             Offenbacher,J., Danzig,J. and Ducar.M.
  TITLE      Creation of genome-wide protein expression libraries using random
             activation of gene expression
  JOURNAL    Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE    21227151
   PUBMED    11329013
COMMENT      Contact: Scott J. Cain
             Athersys, Inc.
             3201 Carnegie Ave, Cleveland, OH 44115, USA
             Tel: 216 431 9900
             Fax: 216 361 9596
             Email: scain@athersys.com
             High quality sequence stop: 505.
FEATURES             Location/Qualifiers
     source          1..849
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."

ORIGIN

Query Match         50.6%;  Score 584.6;  DB 4;  Length 849;
Best Local Similarity   89.3%;  Pred. No. 2.6e-112;
Matches 723;  Conservative  0;  Mismatches 20;  Indels 67;  Gaps  6;

Qy    263 CACTTGAACAAGCAAAAAGTGGAATTGGATTTGGTTTCTGCATACACATTAAATTCAATGTTTGGG 322
          |||| |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    '10 CCCTTGAACAAGCAAAAAGTGGATTTGGTTTCTGCATACACATTAAATTCAAATGTTTTGGG 69

Qy    323 TTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACACATCCAGTAAAACAGGAAATTGGAAA 382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 TTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAACACATCCAGTAAAACAGGAAATTGGAAA 129

Qy    383 GAATCAGGTATATATGAACAGAGTCAAGGAAATAACAGACAAGAAAAAGGCTGGCAAGC 442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 GAATCAGAGTATATGAACAGAGTCAAGGAAAATAACAGACAAGAAAAAGGCTGGCAAGC 189

Qy    443 TGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGAACCAAAAATCGAAAA 502
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 TGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAATGCCCTCTGGGAACCAAAAATCGAAAA 249
```

```
RESULT 9
BI753996
LOCUS        BI753996              878 bp    mRNA            EST 25-SEP-2001
DEFINITION   603027431F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5197999 5',
             mRNA sequence.
ACCESSION    BI753996
VERSION      BI753996.1  GI:15745574
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 878)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM11495  row: m  column: 08
             High quality sequence stop: 843.
FEATURES             Location/Qualifiers
     source          1..878
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
```

```
Qy    503 ATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAAAAGTAACTTTTTGGTTTTTGATGTAC 562
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 ATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAAAAGTAACTTTTTGGTTTTTGATGTAC 309

Qy    563 ACATATTCAAAAAGTAC------------------------------------ 579
          |||||||||||||||||
Db    310 ACATATTCAAAAAGTACATCTTCCCCCCCCCCCCGAAAATAATTAATTCTGTGGCAGGGCAAG 369

Qy    580 ----------------ATTAATATGTTAATCACA---GTAATATGTAAAGCTAAAATAC-T 618
                          |||||| |||||||         || |||||||| ||||||| |
Db    370 GTTTAAATGTGTTTCTTATTTAATATGTAAATTCACAGTAATATGTTAAAGCTAAAATACTT 429

Qy    619 TCCTCTCCAAAGATACATTATCTTTATTGATTAGCACTGAGGATTTTAACATTGTGATATA 678
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 TCCTCTCCAAAGATACATTATCTTTATTGATTAGCACTGAGGATTTTAACATTGTGATATA 489

Qy    679 TTATATATTTATTAATTTACCATTCTCTTGATGAGACTCTTATTTCTTTATATAGGTCAGTC 738
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 TTATATATTTATTAATTTACCATTCTCTTGATGAGACTCTTATTTCTTTATATAGGTCAGTC 549

Qy    739 TTGCAAGTACCATTTTATTAAGCAGCTGTGAAATTTAAGTGAAAGTTCTTTGTTAAACATT 798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 TTGCAAGTACCATTTTATTAAGCAGCTGTGAAATTTAAGTGAAAGTTCTTTGTTAAACATT 609

Qy    799 TGTACTATTTTAAATGAATAAATGACCTTATGAAGTATGCCTATCGTAGGCTGAAATTATA 858
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610 TGTACTATTTTAAATGAATAAATGACCTTATGAAGTATGCCTATCGTAGGCTGAAATTATA 669

Qy    859 GGTACATCTGTTTTCACTA--TATGATATTAAGAAAGGCTG--AATGACTTAAATGTTCATT 916
          ||||||||||||||||||   |||||||||||||||||||  |||||||||||||||||||
Db    670 GGTACATCTGTTTTCACTATTTATGATATTAAGAAAGGCGTGAAATGACTTAAATGTTCATT 729

Qy    917 TTTTTTCTGTATAGATACTTTATTCATGTTTTTCATGATTTTTAGGGAATTACTCGCTTGTTGAT 976
          |||| ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    730 TTNTTCTGTATAGATACCTTATCAACGTTTTTCATGATTTTTAGGGAATTAACTGCTTGTTGAT 789

Qy    977 ATTCAAAGTGTGAAACTAAAAGTTTATGGT 1006
          ||||| |||||||||||||||||||||||
Db    790 A-TCAAAGGGTGAAACTAACACTTATGGGT 818
```

```
Db    811 AATGACCTTATGAAAGTATGCTATCTGTAGGCTGGAATTATAGGGTACATCTGTTTTCAC 870
          |||||||||
Qy    876 TATATGA 882
          |||||||
Db    871 TATATGA 877
```

```
ORIGIN
LOCUS       BU854138              781 bp   mRNA   linear   EST 16-OCT-2002
DEFINITION  AGENCOURT_10403170 NIH_MCC_82 Homo sapiens cDNA clone IMAGE:6621305
            5', mRNA sequence.
ACCESSION   BU854138
VERSION     BU854138.1  GI:24039104
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 781)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2872 row: m column: 17
            High quality sequence stop: 554.
FEATURES             Location/Qualifiers
     source          1..781
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6621305"
                     /clone_lib="NIH_MCC_82"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1:
                     SfiI (ggccgcctcggcc); Site 2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3', and 3' adaptor sequence:
                     5'-ATTCTAGAGCGCGAGCCCGACATG-dT(30)BN-3', (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size
                     1.35 Kb (range 0.9-4.0 Kb). 14/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech Laboratories (Palo
                     Alto, CA)."
```

```
Query Match
Best Local Similarity  50.0%;   Score 578.2; DB 5; Length 781;
Matches 686; Conservative 0; Mismatches 14; Indels 70; Gaps 3;
                                Pred. No. 5.7e-111;

Qy    193 GGTGCTGTGCGATGAGATGCTGAAGACCATGATGATGTCTGTTTCTGAAAATGAGTTGTTGTTGCAG 252
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 GGTGCCGTGGCGATGAGATGCTGAAGACCATGATGATGTCGTGTTTCTAGAAATGAGTTGTTGTTGCAG 62

Qy    253 AAGTTGGATCCACTTGAACAAGCAAAAAGTGGATTGGATTGGTTTTGCATACACATTAAATTCA 312
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 AAGTTGGATCCACTTGAACAAGCAAAAAGTGGATTGGTTTTGCATACACATTAAATTCA 122

Qy    313 ATGTTTGGGTTTTATTTGGGCAACCCAAGGAGTTAATCCTAAGGAACATCAGTAAAACAG 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 ATGTTTGGGTTTTATTTGGGCAACCCAAGGAGTTAATCCTAAGGAACATCCAGTAAAACAG 182

Qy    373 GAATTGGAAAAGAATCAGAGTATATATGAACAGAGTCAAGGAGAAATAACAGCAAGAAAAAAG 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 GAATTGGAAAAGAATCAGAGTATATATGAACAGAGTCAAGGAGAAATAACAGCAAGAAAAG 242
```

```
Query Match
Best Local Similarity  50.1%;   Score 579.6; DB 4; Length 878;
Matches 760; Conservative 0; Mismatches 15; Indels 72; Gaps 10;
                                Pred. No. 2.9e-111;

Qy    107 AGCTGGGCCCATAATGAGCAGGTGAGAAATAAATTAATGAGACAGACTATCCAGTAGAGAAATTCACGAGT 166
          || |||||||||||||| ||||||||| || |||||||||||||||||||||||||||||||||||
Db     32 AGTCAGCCCATAATGGCAGGTGAAGAAATTAATGAACAGACTATCCAGTAGAGAAATTCACGAGT 91

Qy    167 ATTTGTCAGCGTTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATGT 226
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 AATTGTCAGCCGTTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATGT 151

Qy    227 CTGTTTCTGAGAAATGAGTTGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAAGTGGATT 286
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 CTGTTTCTAGAAATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAAGTGGACT 211

Qy    287 TGGTTTTGCATACACATTAAATTCAATGTTTGGGTTTTATTTGGGCAACCCAAGGAGTTA 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 TGGTTTTGCATACACATTAAATTCAATGTTTGGGTTTTATTTGGGCAACCCAAGGAGTTA 271

Qy    347 ATCCTAAGGAACATCCAGTAAAACAGGAATTGGAAAAGAATCAGAGTATATATGAACAGAG 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    272 ATCCTAAGGAACATCCAGTAAAACAGGAATTGGAAAAGAATCAGAGTATATATGAACAGAG 331

Qy    407 TCAAGGAAATAACAGCAAGAAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGAT 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    332 TCAAGGAAATAACAGCAAGAAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGAT 391

Qy    467 TTGTAAAAAATGCCCTCTGGGAACCAAAATGCATCAAAAGTTGCCAATAA-A 525
          ||||||||||||||||||||||||| ||||||||  ||||||||| |
Db    392 TTGTAAAAAATGCCCTCTGGGAACCAAAAACCGAAAAATGCATCAAAAGTTGCCAATAACA 451

Qy    526 GGAAAAGTAAAAGTAACTTAACTTTTTTGGGTTTTGATTGTACACATATTCAAAAAGTAC- 579
          |||||||||||||||| |||||||||| ||||| |||||||||||||||||||||
Db    452 GGACAAAGTAAAAGTAAACTTTTNTGGTTTTGATTGTACACATATTCAAAAAGTACATCTTC 511

Qy    580 ----------ATTAA 584
                    |||||
Db    512 CCCCAGGCCCCGCAAAATAAATAATTCTGTGGCAAGGTTTAAAATGTGTACTTATTATTAA 571

Qy    585 TATGTGATAATCACA---GTAATATGTAAAGCTAAATAC-TTCCCTCTCCAAAGATCATTATC- 639
          ||||||||| || |           || || ||||||| ||||||||| |||||||||||||
Db    572 TATGTAAATTCACAGTAAAATATGTAAAAGCTAAATACTTCCCTCTCCAAAGATCATTATCG 631

Qy    640 TTTAATGATTATTAGCACTGACTGAAGGA-TTTTAACATTGTGATATATATATATAATTTACC 698
          |||||||||||| |  |  ||||| |||| ||||||||||||| || ||| || |||||||
Db    632 TTTATTGAAATTCACAGTAAAATGTAAAAGCTAAATACTTCCTCTCCAAAGATCATTATCG 691

Qy    699 ATCTCTTGATGATGAGACTCTTATTTCTTTATATAGGTCAGTACACATTAAATTACC 758
          |||| |||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    692 ATCTC-TGATGAGACTCTTATTTCTTTATATAGGTCAGTACACATTTTATAA 750

Qy    759 GCAGCTGTGAAATTAAGTGAAAGTTCTT-TGTAAACATTTGTTGTACTATTTTTAAATGAAT 817
          |||||||||||||| ||||| |||||| |||||||||||| |||||  ||||||||||||
Db    751 GCAGCTGTGAAATTAAGTGAAAGTTCTGTTGTACTATCTATTTTAAATGAAT 810

Qy    818 AATGACCTTATG-AAGTATGCTATCTGTAGGCTGGAATTATA-GGTACATCTGTTTTCAC 875
          ||||||||| ||  |||||| ||||||||||||||||||||  |||||||||||||||||
Db    811 AATGACCTTATGAAAGTATGCTATCTGTAGGCTGGAATTATAGGGTACATCTGTTTTCAC 870
```

```
ORIGIN
LOCUS       BU854138              781 bp   mRNA   linear   EST 16-OCT-2002
DEFINITION  AGENCOURT_10403170 NIH_MCC_82 Homo sapiens cDNA clone IMAGE:6621305
            5', mRNA sequence.

RESULT 10
```

```
Db    876 TATATGA 882
          |||||||
Qy    871 TATATGA 877
```

```
Query Match
Best Local Similarity  50.1%;   89.7%;
Matches 760; Conservative 0; Mismatches 15; Indels 72; Gaps 10;

ORIGIN
/db_xref="taxon:9606"
/clone="IMAGE:5197999"
/lab_host="DH10B"
/clone_lib="NIH_MGC_114"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI
Site 2: EcoRV (destroyed); RNA source anonymous pool of 6
male_brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed for
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."
```

```
RESULT 11
LOCUS       AL540005/c
DEFINITION  AL540005 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF031YH10 3-PRIME, mRNA sequence.
ACCESSION   AL540005
VERSION     AL540005.3  GI:45715641
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 887)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:31264568.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 1461.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DF031DD05NPi&c=1461.r.
FEATURES             Location/Qualifiers
     source          1..887
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF031YH10"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN
```

Query Match   49.8%; Score 575.4; DB 1; Length 887;
Best Local Similarity   78.4%; Pred. No. 2.2e-110;
Matches 695; Conservative 65; Mismatches 51; Indels 75; Gaps 5;

```
Qy    433 GCTGGCCAAGCTGGACAGAGAGCTGCAGCTTCAAGATTTGTAAAAAATGCCTCTGGGAACCA 492
          ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    243 GCTGGCCAAGCTGGACAGAGAGCTGCAGCTTCAAGATTTGTAAAAAATGCCTCTGGGAACCA 302

Qy    493 AAATCGAAAAATGCATCAAAAGTTGCCCAATAAAGGAAAAAGTAAAAGTTAACTTTTTGGT 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 AAATCGAAAAATGCATCAAAAGTTGCCCAATAAAGGAAAAAGTAAAAGTTAACTTTTTGGT 362

Qy    553 TTTGATGTACACATATTCAAAAAGTAC----------------------------- 579
          |||||||| |||||||||||||||||
Db    363 TTTGATGTACACATATTCAAAAAGTACATCTTCCCCCCCCCCCCCAAATAT------ 422

Qy    580 ---------------ATTAATATGTAA----TCACAGTAATAT 603
                         |||||||||||||      |||||||||||
Db    423 TCTGTGGCAGGGCAAGGGTTTAAATGTGTTTCTTATTAATATGTAAATTTACACAGTAAATAT 482

Qy    604 GTAAAGCTAAATAC-TTCCTCTCCAAAGATCATTATCTTTATTAATGTAGATTAGCACTGAGGATT 662
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 GTAAAGCTAAATACTTTCCTCTCCAAAGATCATTATCTTTATTAATGTAGATTAGCACTGAGGATT 542

Qy    663 TTAACATTGTGATATATATATATATATATATATATACCATCTCTCTGATTGAGACTCTTATTTC 722
          |||||||| |||||| |||||||||||||||||||||
Db    543 TTAACATTGTGGATATATTATATATTTACCATCTCTCTTGATTGAGACTCTTATTTC 602

Qy    723 TTTATATAGGTCAGTCTTGCCAAGTACCAATTTATATAAGCAGCTGTGAAATTTAAGTGAAAT 782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 TTTATATAGGTCAGTCTTGCCAAGTACCAATTTATATAAGCAGCTGTGAAATTTAAGTGAAAT 662

Qy    783 GTTCTTTGTAAAACATTTGTACTATTTTAAATGAATGACCTTAAGTAGTATGCTATCT 842
          |||||||||||||||| |||||||||||||||||||||||||||||
Db    663 GTTCTTTGTAAACATTTGTACTATTTTAAAATGAATGACCTTAAGTAGTATGCTATCT 722

Qy    843 GTAGGCTGTAATTATAAGTACATCTGTTTTCACTATATATGATATTAAGCTA 892
          |||||||||||||||||||||||||||||||||||||||||||||
Db    723 GTAGGCTGTAATTATAAGTACATCTGTTTTCACTATATATGATATTAAGAA 772
```

Query Match (continued — second alignment block):

```
Qy    324 TTATTTCGGCAACCCAAGGAGGTTAATCCTAAGGACACTGAGGATTTTAACATTG 383
          || |||| :|||   |||  ||:|:|||
Db    526 AATACTTTCCCCCCCCCAAAGATCATTATCTTTATTAATGTCTTATTAATGTAAAGCAGGGAATTGGAAAG 467

Qy    384 AATCAGAGTATATATGAACAGAGTCAAGGAAATACAGACAAGACAAGAAAAAGGCTAAGACGGT 443
          |||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db    384 TTTTTTTTKGCACCCCARGGSKTWATCCTARGGACAYCCCCTAAAAACSGGRWTKGGRAAG 827

Qy    444 GGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAATGCCCTCTGGGAACCAAAATCGAAAAA 503
          ||||:|||| || :|||||||||||| ||||||||||WDCCCYTGGGRACCAAAAWYGAAAAA
Db    766 GGRMRGGGGGGSSVGCTTCAAGATTTTTAAAAAAWDCCCYTGGGRACCAAAAWYGAAAAA 707

Qy    504 TGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAAAGTTAACTTTTTGGTTTTGATGTACA 563
          |||:|||||||||||| |||||||||:|| ||||||||||||| |||||||||||||
Db    706 AGVAWAAAAAARTTGCCAAAAAAAGGRAATAAMARAMAWAAATTTTKGTTTTKATKYAAA 647

Qy    564 CATATTCAAAAAGTAC--------------- 579
          |:||||:|||||:||:
Db    564 CATATTCAAAAAGTAC----------------------- 579

Qy    580 ----------ATTAATATGTAAATCACA---GTAATATGTAAAGCTA 612
                    |||||||||||| :||   |||| |:|||||||||CHA
Db    646 AAWATTMAAAAART AMATYTTCCCCCCCCCCCGCACAAAATAATTCTGTGGCA 587
```

```
Qy   1089 TTGTATGTTGATGGATGAATTTTGGCATGATGACTGTACTCTCAAT 1134
          ||||||||||||||||||||||||||||||||||||||||||||||
Db     46 TTGTATGTTGATGGATGAATTTTGGCATGATGACTGTACTCTCAAT 1
```

RESULT 12
```
LOCUS       BM679939                807 bp    mRNA    linear   EST 27-FEB-2002
DEFINITION  UI-E-EOI-aix-a-14-0-UI.s1 UI-E-EOI Homo sapiens cDNA clone
            UI-E-EOI-aix-a-14-0-UI 3', mRNA sequence.
ACCESSION   BM679939
VERSION     BM679939.1  GI:18989835
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 807)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
PUBMED      8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 476-498, >AT rich#low_complexity (matched compliment)
            Seq primer: M13 Forward
            PolyA=Yes.
FEATURES             Location/Qualifiers
     source          1..807
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EOI-aix-a-14-0-UI"
                     /tissue_type="fetal eye"
                     /dev_stage="fetal"
                     /lab_host="DH10B (life Technologies)"
                     /clone_lib="UI-E-EOI"
                     /note="Organ: eye; Vector: pT773-Pac (Pharmacia) with a
                     modified polylinker; Site 1: EcoR I; Site 2: Not I;
                     UI-E-EOI is a normalized cDNA library containing the
                     following tissue(s): fetal eye. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT773-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     CGCGTATACC. This library was created for the program, Gene
                     Discovery in the Visual System, supported by National Eye
                     Institute (NEI)."
                     TAG_TISSUE=human fetal eye
                     TAG_LIB=UI-E-EOI
                     TAG_SEQ=CCCGTATACC"
ORIGIN
```

Query Match   49.1%; Score 568; DB 4; Length 807;
Best Local Similarity   89.8%; Pred. No. 7.9e-109;
Matches 692; Conservative   0; Mismatches 11; Indels 68; Gaps 5;

```
Qy        452 GTGGCAGCTTCAAGATTGTTGTAAAAAATGCCTCT--GGGGAACCAAAATCGAAAAATGCATCAA 511
              |||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db        783 GTGGCAGCTTCAAGATT--GTAAAAAATGCCCTCTGGGAACCAAAAACCGAAAAATGCATCAA 726

Qy        512 AAGTTGCCAATAAAGGAAAAGTAAAAAGTTAACTTTTTTGGGTTTTGATGTGTACACATATTCA 571
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        725 AAGTTGCCAATAAAGGAAAAGTAAAAAGTTAACTTTTTTGGGTTNTGATGTGTACACATATTCA 666

Qy        572 AAAAGTAC--------------------------------------------------- 579
              ||||||||
Db        665 AAAAGTACATCTTCCCCACCCCCCCCCCGAAAAATAATTCTGTGGCAGGCGAAGGTTTAAAT 606

Qy        580 ------ATTAATATGTAATCACA---GTAATATGTAAAGCTAAATAC-TTCCTCTCC 626
                    |||||||||||||||   |||||||||||||||||||||| |||||||||
Db        605 GTGTTCTTATTATTAATATGTAATTCACAGTAATATGTAAAGCTAAATACTTCCTCTCC 546

Qy        627 AAAGATCATTATCTTTATTGATTAACTACACTGAGGATTTTAACATTGTGATATTGATATTGATTAATAT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 AAAGATCATTATCTTTATTGATTAACTACACTGAGGATTTTAACATTGTGATATTGATATTGATTAATAT 486

Qy        687 TTATAATTACCATCTCTTGATGAGACTCTTATTCTTTTATATATAGGTCAGTCGTTGCAAGT 746
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 TTATAATTTACCATCTCTTGATGAGACTCTTATTCTTTTATATATAGGTCAGTCGTTGCAAGT 426

Qy        747 ACCATTTTATAAGAGCAGCTGGTGAAAATTTAAGTGAAAGTTCTTTGTAAACAATTGTACTAT 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 ACCATTTTATAAGAGCAGCTGGTGAAAATTTAAGTGAAAGTTCTTTGTAAACAATTGTACTAT 366

Qy        807 TTTAAATGAATAATGACCTTAGTAAGTATCTGGCATGTTGCCTCGTTTGTTGGTACTACATC 866
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 TTTAAATGAATAATGACCTTAGTAAGTATCTGGCATGTTGCCTCGTTTGTTGGTACTACATC 306

Qy        867 TGTTTTCACTATATGATTATTAAGAAAGCGTG--AATGACTTAAAAGTTCATTTTTTTCTGT 925
              |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db        305 TGTTTTCACTATATGATTATTAAGAAAGCGTGAAATGACTTAAAAGTTCATTTTTTTCTGT 246

Qy        926 ATAGATACTTTATCAGTTTTTCATGAATTTAGGAATTACTGCTTTTGTTGATATTCAAAGT 985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 ATAGATACTTTATCAGTTTTTCATGAATTTAGGAATTACTGCTTTTGTTGATATTCAAAGT 186

Qy        986 GTGAAACTAAAAGTTTATGGTTGTGTTTACTTTAATTCTTGGCATGTTGCCTCTATGTCCCATT 1045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 GTGAAACTAAAAGTTTATGGTTGTGTTTACTTTAATTCTTGGCATGTTGCCTCTATGTCCCATT 126

Qy       1046 TAAAATAAAATACATTCTCATTAACTTTAGATGGGAAATAAAGGTTGTATGTTGATGGATG 1105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 TAAAATAAAATACATTCTCATTAACTTTAGATGGGAAATAAAGGTTGTATGTTGATGGATG 66

Qy       1106 AATTTTGGCATGATGACTGTACTCTCAATAAAGGCTGAAAAATGTTGTAAA 1156
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AATTTTGGCATGATGACTGTACTCTCAATAAAGGCTGAAAAATGTTGTAPA 15
```

RESULT 13
```
LOCUS       AI800532                909 bp    mRNA    linear   EST 13-DEC-1999
DEFINITION  tc12a11.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:2063612
            3', similar to TR:Q13901 Q13901 C1D DNA-BINDING PROTEIN. ;, mRNA
            sequence.
ACCESSION   AI800532
VERSION     AI800532.1  GI:5366004
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 909)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1095   Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 407.

FEATURES   Location/Qualifiers
source   1..909
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="IMAGE:2063612"
          /tissue_type="Pooled human melanocyte, fetal heart, and
          pregnant uterus"
          /clone_lib="Soares NhHMPu_S1"
          /lab_host="DH10B"
          /note="Organ: mixed (see below); Vector: pT7T3D-Pac
          (Pharmacia) with a modified polylinker; Site_1: Not I;
          Site_2: Eco RI; Equal amounts of plasmid DNA from three
          normalized libraries (melanocyte 2NbHM, pregnant uterus
          NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
          were made in vitro. Following HAP purification, this DNA
          was used as tracer in a subtractive hybridization
          reaction. The driver was PCR-amplified cDNAs from pools of
          5,000 clones made from the same 3 libraries. The pools
          consisted of I.M.A.G.E. clones 260232-265223,
          340488-345479, and 484488-489479."

ORIGIN

Query Match   49.0%;   Score 567; DB 1; Length 909;
Best Local Similarity   84.3%;   Pred. No. 1.3e-108;
Matches 758; Conservative   0; Mismatches   72; Indels   69; Gaps   8;

Qy   325 TATTTGGCAACCCAAGGAGTTAATCCTAAGGACAACATCCCAGTAAAAACAGGAATTGGAAAGA 384
          ||||||||||||||||||||| || ||||||||||||||||||||||||||||||||||||||
Db   909 TTTTTTTGCCAACCAGGAGTTAATCTTAAGGACCAT-CAGTTAAAACCGGAATTGGAAAAG 851

Qy   385 ATCAGAGTATATATGAAACAGAGTCAAGGAAATACAGACAACAAGAAAAGGCT-GGCAAGCT 443
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db   850 AATCAAGGTTTTATTGACCAAGTCAAGGAAATACCAAGGCTGGGCAAGTT 791

Qy   444 -GGACAGAGGTGCAGCTTCAAGA--TTTGTAAAAATGCCCTCTGGGGAACCAAAAATCGAAAA 502
          ||||| |||||| ||||||||| ||||||||||||||||||||||||| |||||||
Db   790 GGCCAGAGGTGCAGCCTTCCAGATTTTGTAAAAATGGCCTTTGGGACCCAAACCCGAAAA 731

Qy   503 ATGCATCAAAAGTTCCCAATAAAAGGAAAAGTTAAACTTTTTTGGTTTTGATGTAC 562
          ||||||||||||||||||||| |||||||||||||| |||| ||||||||||||
Db   730 ATGCATCAAAAGTGGCCATAAAAGAAAAAGTANAAGTTAACTTTTTTTAATGGAA 671

Qy   563 ACATATTCAAAAAGTAC-------------------------------------- 579
          ||||| |||||||||||
Db   670 CAAATTCAAAAAGGTACATCCCCCCCCCCCCCCAAATAATTCTGTGTCGGCAGGGCAAG 611

Qy   580 ----------------ATTAATAGTAATCACA--GTAATAGTGAAAGCTAAATAC-T 618
          ||||||||||| || |||||| |
Db   670 CAAATTCAAAAAGGTACATCCCCCCCCCCCCCCAAATAATTCTGTGTCGGCAGGGCAAG 611

Qy   610 GTTTAAAATGTGTTTCTTATTAATATAGTAAATTCACAGTAAATATAGTAAAGCTAAATACTT 551
          |||||||||||||||||||||||| ||

Qy   619 TCCTCTCCAAAGAYCATTANTCTTTATTAGTTAGCACTGAGGATTTTAACATTTGCATANTA 678
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   550 TCCTCTCCAAAGAYCATTATCTTTATTAGTTAGCACTGAGGATTTTTAACATNTGCATANTA 491

Qy   679 TTTATATATTATTAATTACCATCTCTTGATGAGAGCTCTTATTTTTCTTTATTATAGGTCAGTC 738
          || |||| ||||| ||||||| |||||||||||||||||||||||||| ||||||||||||
Db   490 -TATATATTTATTAATTTACAATCTCCTTGATGAGACTCTAATTCCTTTATTATAGGTCAGTC 432

Qy   739 TTGCAAGTACCCATTTATTAPAGCAGCTGTGAAAATTTAAGTGAAAATTTCTTTGTAAACATT 798
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   431 TTGGAAGTACCCATTTATTAAGCAGCTGTGAAAATTTAAGTGAAAATGTTCTTTGTAAACATT 372

Qy   799 TGTACTATTTTAAATGAATAATGCATTATTATGAAGTATAGTTAGGGCTGAAATTATA 858
          ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   371 TGTACTATTTTAAATGAAATAAATGACCTTATGCTATCCGTAGGGCTGAAATTATA 312

Qy   859 GGTACATCTGTTTCACTATATGATATATGATTATTAAGAAAGGCGTG-AATGACTTAAATGTTCATTT 917
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   311 GGTACATCTGTTTCACTATATGATATATGATTATTAAGAAAGGCGTGAAATGACTTAAATGTTCATTT 252

RESULT 14
BG213559
BG213559
LOCUS   BG213559   1209 bp   mRNA   linear   EST 21-APR-2001
DEFINITION RST33166 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG213559
VERSION BG213559.1   GI:13735246
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1209)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cochren,K., Lo,K.,
Offenbacher,J., Danzig,J. and Ducar,M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21271151
PUBMED 11329013
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 566.

FEATURES   Location/Qualifiers
source   1..1209
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /cell_line="HT1080"
          /clone_lib="Athersys RAGE Library"
          /note="See 'Creation of Genome-wide Protein Expression
          Libraries using Random Activation of Gene Expression',
          Nature Biotechnology, in press. Note that even though the
          cell type indicated is HT1080, since a random activation
          method was used, these sequence tags are not necessarily
          expressed in HT1080 under normal circumstances."

ORIGIN

Query Match   48.9%;   Score 564.8; DB 4; Length 1209;
Best Local Similarity   86.1%;   Pred. No. 3.8e-108;
Matches 749; Conservative   0; Mismatches   46; Indels   75; Gaps   9;

Db   311 GGTACATCTGTTTCACTATATGATATATGATTATTAAGAAAGGCGTGAAATGACTTAAATGTTCATTT 252
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||

Db   311 GGTACATCTGTTTCACTATATGATATATGATTATTAAGAAAGGCGTGAAATGACTTAAATGTTCATTT 252

Qy   918 TTTTCTGTATAGATACTTTTATCATGTTTTCATGATTTTAGGAATTACTGCTTTTGTTGATA 977
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||

Db   251 TTTTCTGTATAGATACTTTTATCATGTTTTCATGATTTTAGGAATTACTGCTTTTGTTGATA 192

Qy   978 TTCAAAGCTGTGAAAACTAAAAAGTTATGGTTGTAGCTTTAAATCTTGGCATGGTGCCTCTAT 1037
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||

Db   191 TTCAAAGCTGTGAAAACTAAAAAGTTATGGTTGTAGCTTTAAATCTTGGCATGGTGCCTCTAT 132

Qy   1038 GTCCCATTTAAAATAAAATAACATTCTCCATTAACTTTAGATGTGGAAATAAGGTGTGTATGT 1097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   131 GTCCCATTTAAAATAAAATAACATTCTCCATTAACTTTAGATGTGGAAATAAGGTGTGTATGTT 72

Qy   1098 GATGGATGAAATTTTGGCATGATGAGTGACTGTACTCTCAATAAAAGGCTGAAAAGTGTTGTAAAA 1156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   71 GATGGATGAAATTTTGGCATGATGAGTGACTGTACTCTCAATAAAAGGCTGAAAAGTGTTGTAPAA 13

Qy   256 TTGGAATCCACTTGAACAAGCAAAAAGTGGATTTGGTTTTCTGCATACACATTAAATTCAATG 315
          |||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||

Db   5 TTGGAACCACTTGNACAAGCAAAA-TGGAATTGGTTTTCTGCATACACATTAAAATTCAATG 63

Qy   316 TTTTGGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGACAACATCCAGTAAAAACAGGAA 375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
RESULT 15
LOCUS       BG206383         859 bp    mRNA    linear    EST 21-APR-2001
DEFINITION  RST25831 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG206383
VERSION     BG206383.1  GI:13728070
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
   MEDLINE   21227151
    PUBMED   11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 513.
FEATURES             Location/Qualifiers
     source          1..859
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
ORIGIN

Query Match     48.7%;  Score 562.4;  DB 4;  Length 859;
Best Local Similarity   88.1%;
Matches  754;  Conservative    0;  Mismatches  26;  Indels  76;  Gaps  10;
```

```
Qy     64 TTTTGGGTTTATTTGGCAACCCAAGGAGGTTAATCCTAGGACATCTAGTAGTAAAACAGGAA 123
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    263 CACTTGAACAAAGCAAAAGTGGATTTGGTTGGTTCTGCATACACACATTAAATTCAATGTTTTGGG 322

Qy    124 TTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAGAAATAACAGACAAGAAAAAAGGCT 183
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 CCCTTGAACAAGCAAAAGTGGATTTGGTTCTGCATACACACATTAAATTCAATGTTTTGGG 69

Qy    184 GGCAGAGCTGGACAGAGGGTCAGCTTCAAGAGTTGTAAAAAATGCCCTCTGGGAACCAAAA 243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 TTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGACATCTAGTAAAACAGGAATTCAATGTTTAAA 129

Qy    244 TCGAAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAAAAGTAAAAGTTAACTTTTTGGTTTT 303
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 TGGACAGAGGTCCAGCTTCAAGAGTTGTGAAAAAATGCCCTCTGGGAACCAAAATCGAAAA 249

Qy    304 GAGTACACATATTCAAAAAAGTAC 303
          |||||||||||||||||||||||
Db    250 ATGCATCAAAAAGTTGCCAATAAAGGAAAAAAAGTAAAAGTTAACTTTTTGGTTTTGATTGTAC 309

Qy    556 GAGTACACATATTCAAAAAAGTAC 579
          |||||||||||||||||||||||
Db    310 ACATATTCAAAAAAGTAC 309

Qy    580 ------------ATTAATATGTTAATCACA---GTAATATGTAAA 608
                      ||||||||||||||||||   ||||||||||||
Db    370 CAAGGTTTAAAATGTGTTTCTTATTTAATATGTAAATTCACAGTAATATGTAAAGCTAAAT 429

Qy    364 GGCAGGGCAAGGTTTAAAATGTGTTTCTTATTTAATATGTAAATTCACAGTAAAATATGTAAA 423
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 ACTTTCCTCTCCAAAAGATCATTAACTTTATTGATTAGCACTGAGGATTTTAACATTGTGA 489

Qy    424 GCTAAAATACTTTCCTCTCCAAAGATCATTAATCTTTATTGATTAGCCACTGAGGATTTTAAT 483
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 TATATTAATATGTTAATTAATTTACCATCTCTTGATGAGAC 549

Qy    668 ATTGTCATATATTAATATAATTTAATTAATTTAATTTACCCATCTCTTGATGAGACTCTTATTCTTTAT 727
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 AGTCTTGCAAGTACCATTTAATATTAAGCAGCTGTGAAAATTTAAGTGAAATGTTCTTTGTAAA 609

Qy    484 ATTGGGATATATTATATATTATAATTTACCACTCTCTTGATGAGACTCTTATTCTTTTAT 543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    675 TATAATATATATTATTTATAATTACCATCTCTTGATGAGACTCTTATTCTTTATATATAGGTC 734

Qy    728 ATAGGTCAGTCTTCGCAAGTACCATTTAATATTAAGCAGCTGTGAAATTTAAGTGAAATGTTCT 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    735 AGTCTTGCAAGTACCATTTAATATAAGCAGCTGTGAAATTTAAGTGAAATGTTCTTTGTAAA 794

Qy    544 ATAGGTCAGTCTTCGCAAGTACCATTTAATATAAGCAGCTGTGAAATTTAAGTGAAATGTTCT 603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 AGTCTTGCAAGTACCATTTAATATTAAGCAGCTGTGAAATTTAAGTGAAATGTTCTTTGTAAA 609

Qy    788 TTGGTAAACATTTGTACTANTTTAAATGAATAATGAATGAATGTAAGTTAGTAATATGTTCAA 847
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    795 CATTTGTACTATTTTAAATGAATAATGAACCTTATGAAGTAATGCTAATCGTGTAGGCTGAAAT 854

Qy    604 TTGTAAACATTTGTACTATTTTAAATGAATAATGAATGACCTTATGAAGTAATGCTAGG 663

Qy    848 CTGAAATTTATAGGTATGACATCTCGTTTTTCACTATGATGATTTAGAA-AGCGTGAATGACTTA 906

Qy    664 CTGANATTTATAGGTACGATCGTTTCCACTATCTGAATGATGAATATAAGAAANGCGTGAATGACTTA 723

Qy    907 AAGTTTCATTTTTTTCTGTATAGATACTTTTATACATTNTNNTCATGNNTCATGAATTTAGGANNNTACTG 966

Qy    724 NAGTTCATTNTTTNTTCTGTAATAGATACTTTATCATGNNTCATGAATTTAGGANNNTACT 783

Qy    967 CTTTGTTGATATTCAAAAGTGTGAACTAAAAGTTTATATGGTTGTACTTTAAATTCTTGGGCAT 1026

Qy    443 TGGACAGAGGTGCAGCTTCAAGAGTTGTAAAAAATGCCCTCTGGGAACCAAAATCGAAAAA 502

Qy    784 GCTTGNTNGATA-TCAAAGTGTG--AACTAAANACATTATNGGT--GNCTTNNAATCNTTGCAT 839

Db    130 GAATCAGAGTATATATGAACAGAGTCAAGGAGAAATAACAGACAAGAAAGAAAAAGGCTGGCAAGC 189

Qy   1027 GTTGCCTCTAGTCCCATTTAAAATAAAAT 1056

Db    383 GAATCAGAGTATATATGAACAGAGTCAAGGAGAAATAACAGACAAGAAAGAAAAAGGCTGGCAAGC 442

Qy    840 GTTG--CTCTAGTCCATTAANNATAAAAACAT 868

Db     70 TTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGACATCTAGTAAAACAGGAATTCAATGTTTAAA 129

Qy    503 ATGCATCAAAAAGTTGCCAATAAAGGAAAAAAAGTAAAAGTTAACTTTTTGGTTTTGATTGTAC 562

Db    323 TTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGACATCTAGTAAAACAGGAATTCAATGTTTGGG 382

Qy    250 ATGCATCAAAAAGTTGCCAATAAAGGAAAAAAAGTAAAAGTTAACTTTTTGGTTTTGATTGTAC 309

Qy    563 ACATATTCAAAAAAGTAC 579

Db    616 AC-TTCCCTCTCCAAAGATCATTATCTTTATTGATTAGCACTGAGGATTTTAACATTGTGA 674

Db    310 ACATATTCAAAAAAGTAC 579

Db    430 ACTTTCCTCTCCAAAAGATCATTAACTTTATTGATTAGCACTGAGGATTTTAACATTGTGA 489
```

```
Db    610  CATTTGTTACTTATTTTAAATGAATAATGACCTTATGAAGTAATGCTATCTGTAAGCTG-AAT 668

Qy    855  TATAGGTACATCTGTTTTCACTATATGAATATTAAGAAAGGGTG-AATGACTTAAATGTTC 913
                ||||||||||||||||||||||||||||| |||||||||| |||||||||||||||||
Db    669  TATAGGTACATCTGTTTTCACTATATGAATATTAAGAAAGGGTGAAAATGACTTAAATGTTC 728

Qy    914  ATTTTTTTTCTGTATAGATACTTTTATCATGTTTTTCATGATTTTAGGAATTACTGCTTTTGTT 973
                |||||||||||||||||||||||||||||||||||||||| |||||||||||| |
Db    729  ATTTTTTTTCTGTATA-AAACTTTATCATGTTTTTCATGAATTTAAGAAATAACTGC--TTGGT 786

Qy    974  GATATTCAAAGTGTGAAAACTAAAAAGTTTATGGTTGTACTTTAATTCTTGGCATGTTGCCT 1033
                |||||||||| ||| |||||||||||||||||| ||||||||||| || ||| |||
Db    787  GATATTCAAA--GTTGAAACTAAACTTTAAGGTTG-ACTTTAATTCTTGGC-TGGTGGGCT 842

Qy   1034  CTATGTCCCATTTAAA 1049
                ||||||||| | |||
Db    843  CTATGTCCATTAAAAA 858
```

This Page Blank (uspto)

OM protein - protein search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: January 4, 2005, 11:20:50 ; Search time 16.5 Seconds
(without alignments)
822.216 Million cell updates/sec

Title: US-09-701-618A-2
Perfect score: 715
Sequence: 1 MAGEEINEDYPVEIHEYLSA..........LWEPKSKNASKVANKGKSKS 141

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:　　283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
```
1:   PIR_79:*
1:   pir1:*
2:   pir2:*
3:   pir3:*
4:   pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	133.5	18.7	133	2	T41115	sun-cor hormone re
2	110	15.4	184	2	S46808	hypothetical prote
3	83	11.6	207	2	S70534	hypothetical prote
4	83	11.6	207	2	S70533	bbk2.10 protein pr
5	82	11.5	698	2	S52696	bbk2.10 protein pr
6	81.5	11.4	596	2	T44348	myosin heavy chain
7	81.5	11.4	904	2	T03806	GTP binding protei
8	81	11.3	444	2	S64912	hypothetical prote
9	81	11.3	530	2	G71157	probable membrane
10	78.5	11.1	250	2	T22093	hypothetical prote
11	78.5	11.0	508	2	S73430	glycerol kinase gl
12	78.5	11.0	1101	2	T31550	hypothetical prote
13	78.5	11.0	1708	2	AE1866	WD-40 repeat prote
14	78	10.9	630	2	A37097	fimbrin - chicken
15	78	10.9	649	2	S41016	hypothetical prote
16	77.5	10.9	2285	2	T12796	probable transglyc
17	77	10.8	466	2	C89820	cysteinyl-tRNA syn
18	77	10.8	760	2	T31556	hypothetical prote
19	76.5	10.7	705	2	A71211	probable phosphori
20	76	10.6	446	2	G72731	probable malate ox
21	76	10.6	656	2	AH2402	cell division prot
22	76	10.6	674	2	A10922	ATP-dependent DNA
23	75.5	10.6	287	2	E82978	conserved hypothet
24	75.5	10.6	295	2	AB2464	tRNA delta-2-isope
25	75.5	10.6	305	2	AF0305	sugar binding prot
26	75.5	10.6	481	2	T14300	hypothetical prote
27	75	10.5	197	2	G72286	hypothetical prote
28	75	10.5	291	2	E69777	transcription regu
29	75	10.5	415	2	A48359	glutamyl-tRNA redu
30	75	10.5	424	2	E98008	conserved hypothet
31	75	10.5	992	2	S54396	protein L precurso
32	74.5	10.4	248	2	D71672	hypothetical prote
33	74.5	10.4	312	2	D84956	hypothetical prote
34	74.5	10.4	710	2	S40934	hypothetical prote
35	74.5	10.4	1021	2	T15900	hypothetical prote
36	74.5	10.4	1311	2	A56390	mannosyl-glycoprot
37	74	10.3	1211	2	T32902	hypothetical prote
38	74	10.3	465	2	D90426	glucan phosphoryla
39	74	10.3	471	2	S76021	probable trigger f
40	74	10.3	478	2	C59095	hypothetical prote
41	74	10.3	560	2	T32288	hypothetical prote
42	74	10.3	629	2	A56536	plastin, intestina
43	74	10.3	661	2	S67177	hypothetical prote
44	73.5	10.3	441	2	B71816	hypothetical prote
45	73.5	10.3	604	2	T19682	hypothetical prote

ALIGNMENTS

RESULT 1
T41115
sun-cor hormone receptor repression complex homolog - fission yeast (Schizosaccharomyces
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T41115
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21969
A;Accession: T41115
A;Molecule type: DNA
A;Status: Preliminary; translated from GB/EMBL/DDBJ
A;Experimental source: strain 972h-; cosmid c1739
A;Cross-references: UNIPROT:O74469; EMBL:AL031540; PIDN:CAA20781.1; GSPDB:GN00068; SPDB:S
C;Genetics:
A;Genetics:
C;Genetics: SPDB:SPCC1739.07
A;Map position: 3

```
Query Match       18.7%;  Score 133.5;  DB 2;  Length 133;
Best Local Similarity   27.8%;
Matches  37;  Conservative  25;  Mismatches  54;  Indels  17;  Gaps   4;

Qy      16 EYLSAFE---NSIGAVDEMLKTMMSV-SRNELLQKLIDPLEQAKVDLVSAYTLNSMFWVYL 71
              ||| ||      ::| || ::  : |  :||| |||:||||: :|||:| |||  :|||
Db       4 EYSELFERLNKQLDNVEDVLKPLKDAESIFELABGKSELEQAKLYTMSYAINSTLYSFY 63

Qy      72 ATQGVNPKEHPVKQELERIRVYMNRV---KEITDKKKAGKLDRGAASRFVKNALWEPKS 127
             :: |||  ::: || :|:|: |::|     |:| ||| |:|| : :::|  || |||
Db      64 KLNGIDASERPVMQELQRVKNYISKIQQAEKNVNPKTEAVN-------TSNAAISSSS 114

Qy     128 KNASKVANKGKSK 140
               |  ||||||||
Db     115 SNRPKVAKDAATR 127
```

RESULT 2
S46808
hypothetical protein YHR081w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jul-2004
C;Accession: S46808
R;Favello, T.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of S. cerevisiae cosmid 9205.
A;Reference number: S46795
A;Accession: S46808
A;Molecule type: DNA
A;Residues: 1-184 <FAV>
A;Cross-references: UNIPROT:P38801; EMBL:U10556; NID:g500825; PID:g500829; GSPDB:GN00008;
C;Genetics:
A;Gene: MIPS:YHR081w
```

A;Cross-references: SGD:S0001123
A;Map position: 8R

Query Match      15.4%;  Score 110; DB 2; Length 184;
Best Local Similarity   25.9%;  Pred. No. 0.027;
Matches  30;  Conservative  31;  Mismatches  47;  Indels  8;  Gaps  4;

```
Qy  13 EIHEYLSAFENSIGAVDEMLKTMMSVSRNELLOKLDPLEQAKVDLLVS--AYTLINSMFWVY 70
          :| :| : |||:||||||||||:||||||||||||||:||:||||:| :: :
Db   6 KIKPVVRSFSKALIDELKPIEKLTSKSLDEQLLILSD-ERAKLELINRYAYVLLSSLMFAN 64

Qy  71 LATQGVNPKEHPVKQELERIRVYMNRVKE-----ITDKKKGAGKLDRGAASRFVKNAL 122
          | : ::| |: ::| |:: | :|| | :||
Db  65 MKVLGVKDMS-PILGELKRVKSYNDKAKQYDNRITKSNEKSQAEQEKAKNIISNVL 119
```

RESULT 3
S70534
bbK2.10 protein precursor - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S70534
R;Akins, D.R.; Porcella, S.F.; Popova, T.G.; Shevchenko, D.; Baker, S.I.; Li, M.; Norgar
Mol. Microbiol. 18, 507-520, 1995
A;Title: Evidence for in vivo but not in vitro expression of a Borrelia burgdorferi oute
A;Reference number: S70531; MUID:96342380; PMID:8748034
A;Accession: S70534
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-207 <AKI>
A;Cross-references: UNIPROT:Q44732; EMBL:U18292; NID:g3309519; PIDN:AAC26095.1; PID:g896
A;Genetics:
C;Gene: bbK2.10
C;Superfamily: outer surface protein F ospF

RESULT 4
S70533
bbK2.10 protein precursor - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S70533
R;Akins, D.R.; Porcella, S.F.; Popova, T.G.; Shevchenko, D.; Baker, S.I.; Li, M.; Norgar
Mol. Microbiol. 18, 507-520, 1995
A;Title: Evidence for in vivo but not in vitro expression of a Borrelia burgdorferi oute
A;Reference number: S70531; MUID:96342380; PMID:8748034
A;Accession: S70533
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-207 <AKI>
A;Cross-references: UNIPROT:Q44733; EMBL:U19105; NID:g896039; PIDN:AAB40878.1; PID:g8960

F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-207/Product: bbK2.10 protein #status predicted <MAT>

Query Match   11.6%; Score 83; DB 2; Length 207;
Best Local Similarity   24.5%; Pred. No. 6.2;
Matches 36; Conservative 26; Mismatches 53; Indels 32; Gaps 5;

```
Qy   2 AGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLOKLDPLEQAKVDLVSAY 61
          :||:| : ||||||||||||:||||||||||||||||:||:||||||||:|:
Db  25 SGEDIKQNVKEKVEGFLEAKKEELVGGLKIKLGSEAYLKVEELMQADRPQVOAEEQV--- 80

Qy  62 TLINSMFWVYLATQGV--NP--KEHPVKQELERIRVYMNRVKEITDKKKGAGKLDRGAASRF 117
          | ||||||:| :: | |: ::| |:: | ||||:| :||
Db  81 -----AQGVFEDPGLKEKGLEEKIEEL------KELKDSSKKTKEDRKKELEE 122

Qy 118 VRQNALWEPK-------SKNASKVANKGK 138
          || :||
Db 123 AKQKLEEFKRQVESVTENTDKVKNQGK 149
```

RESULT 5
S52696
myosin heavy chain - rainbow trout (fragment)
C;Species: Oncorhynchus mykiss (rainbow trout)
C;Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S52696
R;Gauvry, L.L.; Fauconneau, B.B.
A;Description: Cloning of a fast skeletal myosin heavy chain expressed both in embryo and
submitted to the EMBL Data Library, March 1995
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-698 <GAU>
A;Reference number: S52696
A;Accession: S52696
A;Cross-references: UNIPROT:Q91208; EMBL:Z48794; NID:g755770; PIDN:CAA88724.1; PID:g755,
A;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: ATP

Query Match   11.5%; Score 82; DB 2; Length 698;
Best Local Similarity   22.7%; Pred. No. 31;
Matches 37; Conservative 20; Mismatches 52; Indels 54; Gaps 5;

```
Qy   1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLOKLDPLEQAKVDLVSA 60
          | |||:|| : :: | : | : :|: :|::
Db 527 MMABELKKEQDTSSHL-----------ERMKKNLEITVKDLQHRLDEAEN------- 565

Qy  61 YTLINSMFWVYLATQG-------VNPKEHPVKQELER-------IRVYMNRVKEIT- 101
          | : | :: ||: : :|||
Db 566 -------LAMKGGKKQLQKLESRVRELETEVEABQRRGVDAVKGVRKYERRVKELTY 615

Qy 102 ----DKKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSK 140
          :| | :| : :|: || || :|
Db 616 QTEEDKKNVGRLQDLVDKLQMKVKAYKRHSEEAEEAAANQHMSK 658
```

RESULT 6
T44348
GTP binding protein (imported) - Clostridium histolyticum
C;Species: Clostridium histolyticum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T44348
R;Matsushita, O.; Jung, C.M.; Katayama, S.; Minami, J.; Takahashi, Y.; Okabe, A.
J. Bacteriol. 181, 923-933, 1999
A;Title: Gene duplication and multiplicity of collagenases in Clostridium histolyticum.
A;Reference number: Z22752; MUID:99121032; PMID:9922257
A;Accession: T44348
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-596 <MAT>
A;Cross-references: UNIPROT:Q92NK5; EMBL:AB014075; NID:g3868863; PIDN:BAA34254.1; PID:g3f
C;Genetics:
A;Gene: hflX

RESULT 7
T03806
hypothetical protein P102 - Mycoplasma hyopneumoniae
C;Species: Mycoplasma hyopneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T03806
Gene 214, 13-23, 1998
R;Hsu, T.; Minion, F.C.
A;Title: Molecular analysis of the P97 cilium adhesin operon of Mycoplasma hyopneumoniae
A;Reference number: Z15097; MUID:98391007; PMID:9729120
A;Accession: T03806
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-904 <HSU>
A;Cross-references: UNIPROT:O50187; EMBL:AF012905; NID:g2654768; PIDN:AAC32527.1; PID:g2
A;Genetic code: SGC3
A;Note: P102
C;Superfamily: Mycoplasma hyopneumoniae hypothetical protein P102

Query Match   11.4%; Score 81.5; DB 2; Length 904;
Best Local Similarity 24.4%; Pred. No. 46;
Matches 41; Conservative 29; Mismatches 43; Indels 55; Gaps 10;

Qy   9 DYPVEIHEYLSAFENSIGAV----------DEMLKTMMSVSRNELLQKLDPL-EQAK--- 54
            :||.||:|||||||||:|: | ||.||||||||||:|||||||| |||
Db   418 DLKVDFNKYQAAVAQMFGTIKAVKEFSMPEDQDAKTL---SSNEIKQRVDRLPFELAKTVT 474

Qy   55 -VDLVSAYTLNSMFWVYLATQG---VNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLD 110
            .|||||||||||:|| :|. |||||||||||||||||||||||||||||||||
Db   475 NLENPSEEVLKS----IYLLNTGKYLVDQDQBKVKQELKTV------IEGLKSKANTQKTE 525

Qy   111 RGAASRFVKNALWEPK---------SKNASKVAN--------KGKSKS 141
            ||||||||||||.|:   :|.:.| :||||
Db   526 ---------KNSPTQPKKPEVSLAKTTENSAKTVKVSTPAEEAKGQSOS 565

RESULT 8
S64912
probable membrane protein YLR080w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein L2365
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
C;Accession: S64912
R;Pohl, T.M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64899
A;Accession: S64912
A;Molecule type: DNA
A;Residues: 1-444 <POH>
A;Cross-references: UNIPROT:Q12396; EMBL:Z73252; NID:g1360442; PID:g1360443; GSPDB:GN000
A;Experimental source: strain S288C
A;Gene: MIPS:YLR080w
A;Cross-references: SGD:S0004070

RESULT 9
G71157
hypothetical protein PH0462 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C;Accession: G71157
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: G71157
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-530 <KAW>
A;Cross-references: UNIPROT:O58213; GB:AP000002; NID:g3236129; PIDN:BAA29548.1; PID:g325
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
A;Gene: PH0462

Query Match   11.3%; Score 81; DB 2; Length 530;
Best Local Similarity 25.2%; Pred. No. 47;
Matches 34; Conservative 29; Mismatches 32; Indels 40; Gaps 8;

Qy   10 YPVEIHEYLSAFENSIGA----------------DEMLKTMMSVSR----NELLQKLDPLEQAKVD 56
            ||:||||||||||||.| ||.|||||||| ||||||||||:||||
Db   401 YP-KIYQEITVFGKSLVALKPSPSGHAHEDLTKEYLSITEKALFNHLSKSLDPSTFSQVR 459

Qy   57 LVSAYTLNSM---------FWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKL 109
            |||||||||| |:|||||||||||||||||||||||||||||||||||||
Db   460 IM----LNSMNEDERIFAMFLFT---SAKPEEAKQLLE--KVHVNELKKTID------ 503

Qy   110 DRGAASRFVKNALWE 124
            ||||||||.|||||
Db   504 ---AFYKFYRDIPWE 515

RESULT 10
T22093
hypothetical protein F42E11.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T22093
R;Lennard, N.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19512
A;Accession: T22093
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-250 <WIL>
A;Cross-references: UNIPROT:Q20334; EMBL:Z66562; PIDN:CAA91466.1; GSPDB:GN00028; CESP:F42
A;Experimental source: clone F42E11
A;Genetics:
A;Gene: CESP:F42E11.4
A;Map position: X

[Right column:]

Query Match   11.4%; Score 81.5; DB 2; Length 596;
Best Local Similarity 23.8%; Pred. No. 28;
Matches 29; Conservative 26; Mismatches 32; Indels 35; Gaps 5;

Qy   15 HEYLSAFENSIGAV--DEMLKTMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLA 72
            ||||||||||||| ||.|||||||||.:||||||||||||||||||||||||||||
Db   437 HDLVESFKSTLEEVVNSNLLHVVDSSSEKLYEQIDAVE--------------MVLS 479

Qy   73 TQGVNPKE------------HPVKQELERIRVYMNRVKEITDKKKAGK-LDRGAASRFVKNAL 122
            ||||||| ||||||||||||||||||||||||||| ||||||||||||
Db   480 QLGVDPKECLLVLNLKIDKLTKEELEDIELKLNKYKLVTISAKTGENLD------LILNAI 533

Qy   123 WE 124
            ::
Db   534 CE 535

RESULT 8 [right column continuation]
F;413-429/Domain: transmembrane #status predicted <TMM>

[RESULT 9 right column:]
Query Match   11.3%; Score 81; DB 2; Length 444;
Best Local Similarity 27.8%; Pred. No. 22;
Matches 32; Conservative 20; Mismatches 43; Indels 20; Gaps 5;

Qy   13 EIHEYLSAFENSIGA--VDEMLKTMMSVSRN--ELLQKLDPLEQAKVDLVSAYTLNSMFW 68
            |:||||||||||| |||.||||||||||| |||||||||||||||||||||||||
Db   318 ELPEKMNKLEGKIMANDIDPLLRKMNKIVENERELIQRLRPL---LDLKKTAISDDSFQ 373

Qy   69 VYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKNALW 123
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374 DFLS---MNANLDRLIKEQEKIR------QDAKLYGKQTKGHDEIFSKISVW 416

A;Map position: 12R
C;Keywords: transmembrane protein

A;Introns: 20/3; 62/3; 105/1; 153/2; 202/3;
C;Superfamily: troponin I

```
  Query Match             11.0%;  Score 78.5;  DB 2;  Length 250;
  Best Local Similarity   23.5%;  Pred. No. 19;
  Matches   28;  Conservative  32;  Mismatches  46;  Indels  13;  Gaps  5;

Qy   24 SIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDL--VSAYTLNSMFWVVLATQGVNPKEH 81
           :::|:|:|  |  |::|:|:::|:|:::::|  |:  ::||  |:|:|||: ||||||
Db   97 NVDSIDDHAK--LEAIYNDLFSRLCNLEEKYDINHITTETTIINQLINIEVNDLRGKF- 153

Qy   82 PVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKNALWEPKSKNA-SKVANKGKS 139
           ||||| | |: | |  ::|:::    ||  |:  |   |: ||  | |  |:|  ||
Db  154 -VKPSLKKVSKYDNKPKKMAEAKKED------GSKNLRNNLKTVKKESVFTQIANKKKS 205
```

RESULT 11
S73430
glycerol kinase glpK - Mycoplasma pneumoniae (strain ATCC 29342)
N;Alternate names: hypothetical protein D09_orf508
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
A;Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A;Accession: S73430
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Molecule type: DNA
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Residues: 1-508 <HIM>
A;Cross-references: UNIPROT:P75064; EMBL:AB000012; GB:U00089; NID:g1673755; PIDN:AAB9575
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
A;Gene: glpK
C;Genetic code: SGC3
C;Superfamily: xylulokinase

RESULT 12
T31550
hypothetical protein Y47D3A.26 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
A;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
R;Matthews, L.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21043
A;Accession: T31550
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-1101 <WIL>
A;Cross-references: UNIPROT:Q9U2C1; EMBL:AL117202; PIDN:CAB57898.1; CESP:Y47D3A.26
A;Experimental source: clone Y47D3A
A;Gene: CESP:Y47D3A.26
A;Introns: 135/2; 169/3; 252/3; 287/2; 320/3; 387/2; 441/3; 637/3; 681/3; 786/3; 842/3,
C;Superfamily: hypothetical protein YJL074c

```
  Query Match             11.0%;  Score 78.5;  DB 2;  Length 1101;
  Best Local Similarity   25.6%;  Pred. No. 42;
  Matches   32;  Conservative  20;  Mismatches  54;  Indels  19;  Gaps  5;

Qy   10 YPVEIHEYLSAFENSI----GA-----VDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSAYT 62
Db  298 WQLENQKPVYALEGSVPVAGALKWLRDSLKVMYSAAESDFYAKLAQKEEQEVVFVPAFT 357

Qy   63 -LNSMFW------VYLIATQGVNPKEHPVKQELERIRVYMN-----RVKEITDKKKAGKLD 110
Db  358 GLGAPYWDASARGAIFGIEANTKREHLVKATLEAIAFQANDLIKAMASDLNSSIKKIKAD 417

Qy  111 RGAAS 115
Db  418 GGACN 422
```

RESULT 13
AE1866
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. (strain PCC 7120)
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
A;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anal
A;Accession: AE1866
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Molecule type: DNA
A;Status: preliminary
A;Residues: 1-1708 <KUR>
A;Cross-references: UNIPROT:Q8YZI2; GB:BA000019; PIDN:BAB72436.1; PID:g1129823; GSPDB:(
A;Experimental source: strain PCC 7120
A;Gene: all0478

```
  Query Match             11.0%;  Score 78.5;  DB 2;  Length 1708;
  Best Local Similarity   23.5%;  Pred. No. 1.7e+02;
  Matches   36;  Conservative  31;  Mismatches  51;  Indels  35;  Gaps  5;

Qy    5 EINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAK------ 54
Db  899 QLIHDYLVDLIRYLQQOESSLQAQLDQLRRKVQOSQSEIARLKSELSQKKQSKLTDTHPQ 958

Qy   55 --VDLVSAYTLNSMFWVVLATQGVNPKEHPVKQELERIRVYMN----DKKAGK 108
Db  959 QGLDLVT----------ELRELRKREELTQLEIEQLRGEL-KEKELTAQLAESQKQQR 1005

Qy  109 LDRGAASRFVKNALWEPKSKNASKVANKGKSKS 141
Db 1006 LSEAKLNRSLKIAL-----TASCLAILGLSVS 1032
```

RESULT 14
A37097
fimbrin - chicken
C;Species: Gallus gallus (chicken)
A;Date: 01-Feb-1991 #sequence_revision 14-Jul-1994 #text_change 09-Jul-2004
R;de Arruda, M.V.; Watson, S.; Lin, C.S.; Leavitt, J.; Matsudaira, P.
J. Cell Biol. 111, 1069-1079, 1990
A;Title: Fimbrin is a homologue of the cytoplasmic phosphoprotein plastin and has domain
A;Reference number: A37097
A;Accession: A37097
A;Molecule type: mRNA
A;Residues: 1-630 <DEA>
A;Cross-references: UNIPROT:P19179; GB:X52562; NID:g62889; PIDN:CAA36796.1; PID:g62890
A;Keywords: actin binding; blocked amino end; calcium binding; calmodulin repeat h
F;11-43/Domain: plastin; alpha-actinin actin-binding domain homology <EF1>
F;51-83/Domain: calmodulin repeat homology <EF2>
F;121-372/Domain: alpha-actinin actin-binding domain homology <ACT1>

F;395-621/Domain: alpha-actinin actin-binding domain homology <ACT2>

```
Query Match             10.9%;  Score 78;  DB 1;  Length 630;
Best Local Similarity   25.7%;  Pred. No. 60;
Matches   29;  Conservative  21;  Mismatches  45;  Indels  18;  Gaps   4;

Qy     31 MLKTMMSVSRNELLQKLDPLEQAKVD---LVSAYTLNSMFWVVLATQGVNPKEHPVKQEL 87
          :  :    ::  : |||: : ::  |      :  ||||:|    ::||||||||:   |
Db      1 MENNVTTISREELEELREAPNKIDIDNSGYVSDYELQDLF------------KEASLPLPG 49

Qy     88 ERIRVYMNRVKEITDKKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSK 140
          :  |   ||| :|||||| ||||||||||||       ||||||||||||   :|
Db     50 YKVREIEKIFAVTDSNKDGKIN---FEEFV-SLIQELKSKDVSKSYRKSINK 98
```

RESULT 15
S41016
hypothetical protein T07C4.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S41016
R;Berks, M.
submitted to the EMBL Data Library, January 1994
A;Reference number: S41014
A;Accession: S41016
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-649 <BER>
A;Cross-references: UNIPROT:Q22290; EMBL:Z29443; NID:g1067051; PID:g443830
C;Genetics:
A;Introns: 6/2; 43/3; 96/1; 146/1; 213/3; 310/2; 473/3; 519/3; 630/3

```
Query Match             10.9%;  Score 78;  DB 2;  Length 649;
Best Local Similarity   28.4%;  Pred. No. 62;
Matches   40;  Conservative  20;  Mismatches  65;  Indels  16;  Gaps   7;

Qy     11 PVEIHEYL-SAFE---NSIGAVDEMLKIMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSM 66
          ||: :   ||||     : ||  :  ::  ||||||||| :||||||||:|:| :|||:|
Db    281 PIVLTESIDAAFEAVIQSVSHADATLSMTQNASEEAKVHATPPLPAERKTMVSAELANGK 340

Qy     67 FWVYLATQ---GVNPKEHPVK---QELERIRVYMNRVKEITDKKKAGKLDRGAASRFVK 119
          | ||||| :   ||||||| | :      ||||  ||| :|||||| ||||||||||||
Db    341 PIKHLIEKFDAGVNFAEHKPKSIYAQVLEIGGSAPRVDEVF---SASKEEHAETE-VT 396

Qy    120 NALWEPKSKNASKVANKGKSK 140
          : :  ||||||||||||||:|
Db    397 NVIFRSTSTHSSIIAN-GKEQ 416
```

Search completed: January 4, 2005, 13:45:13
Job time : 23.5 secs

This Page Blank (uspto)

OM protein - protein search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: January 4, 2005, 08:33:30 ; Search time 63.5 Seconds (without alignments)
1277.604 Million cell updates/sec

Title: US-09-701-618A-2
Perfect score: 715
Sequence: 1 MAGEEINEDYPVEIHEYLSA.........LWEPKSKNASKVANKGKSKS 141

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	715	100.0	141	2	Q13901	homo sapien
2	654	91.5	141	2	Q35473	m small uni
3	652	91.2	141	2	Q7TSU0	cricetulus
4	646	90.3	141	2	Q61368	mus musculu
5	549	76.8	111	2	Q7L8E1	homo sapien
6	186.5	26.1	141	2	Q86F82	schistosoma
7	161	22.5	204	2	Q6BJZ2	debaryomyce
8	150.5	21.0	217	2	Q93VA9	arabidopsis
9	139.5	19.5	169	2	Q7QIZ7	anopheles g
10	133.5	18.7	133	2	Q74469	schizosacch
11	133	18.6	194	2	Q6C2H9	yarrowia li
12	132	18.5	207	2	Q6YX09	oryza sativ
13	132	18.5	207	2	BAC99871	oryza sat
14	128.5	18.0	133	2	Q9N3J4	caenorhabdi
15	126.5	17.7	159	2	Q9VXL4	drosophila
16	121	16.9	450	2	Q7SER2	neurospora
17	119	16.6	149	2	Q8I118	plasmodium
18	110	15.4	184	1	P3B801 / YHOL_YEAST	saccharomyc
19	109	15.2	149	2	Q6FUI1	candida gla
20	106.5	14.9	177	2	Q6CMX3	kluyveromyc
21	94.5	13.2	654	2	Q8L3P4	arabidopsis
22	90	12.6	269	2	Q8TZQ4	pyrococcus
23	89	12.4	398	2	Q8IKD8	plasmodium
24	87.5	12.2	915	2	Q32863	mycoplasma
25	86	12.0	1454	2	Q6CT15	kluyveromyc
26	85.5	12.0	244	2	Q9HGN2	schizosacch
27	85.5	11.9	150	2	Q83QG2	enterococcu
28	85	11.9	336	2	Q7RXX8	neurospora
29	85	11.9	1208	2	Q9VXU1	drosophila
30	85	11.9	1398	2	Q9VXU2	drosophila
31	84.5	11.8	530	2	Q6MDN2	parachlamyd
32	84.5	11.8	530	2	CAF23317	parachlam
33	84	11.7	2867	1	Q00799	plasmodium
34	83.5	11.7	129	2	Q977N0	uncultured
35	83.5		2511	2	Q8IL44	plasmodium
36	83	11.6	207	2	Q44732	borrelia bu
37	83	11.6	207	2	Q44733	borrelia bu
38	83	11.6	1249	2	Q8DT19	clostridium
39	83	11.6	2104	2	Q8XM24	streptococc
40	83	11.6	642	2	Q99P70	mus musculu
41	82.5	11.6	642	2	Q9D0F1	mus musculu
42	82.5	11.5	1104	2	Q6BH46	debaryomyce
43	82	11.5	395	2	Q6CQF7	kluyveromyc
44	82	11.5	455	2	Q9SG79	arabidopsis
45	82	11.5	698	2	Q91208	oncorhynchu

ALIGNMENTS

RESULT 1
Q13901

```
ID   Q13901       PRELIMINARY;      PRT;    141 AA.
AC   Q13901;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   CID protein (Nuclear DNA-binding protein).
GN   Name=CID;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Term placenta;
RX   MEDLINE=98136194; PubMed=9459821;
RA   Nehls P., Kecks T., Greferath R., Spiess E., Glaser T., Rothbarth K.,
RA   Stammer H., Werner D.;
RT   "cDNA cloning, recombinant expression and characterization of
RT   polypeptides with exceptional DNA affinity.";
RL   Nucleic Acids Res. 26:1160-1166(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow, Lung, and Uterus;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
```

RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X95592; CAA64845.1; -.
DR   EMBL; BC005584; AAH05584.1; -.
DR   EMBL; BC005584; AAH05584.1; -.
DR   EMBL; BC009589; AAH09589.1; -.
DR   EMBL; BC016284; AAH16284.1; -.
DR   EMBL; BC016284; AAH16284.1; -.
DR   EMBL; BC005235; AAH05235.1; -.
DR   EMBL; BC005235; AAH05235.1; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0003677; F:DNA binding; TAS.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   DNA-binding.
SQ   SEQUENCE 141 AA; 16019 MW; 9976A3BBD5620D63 CRC64;

Query Match   100.0%; Score 715; DB 2; Length 141;
Best Local Similarity   100.0%;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120

Qy  121 ALWEPKSKNASKVANKGKSKS 141
          |||||||||||||||||||||
Db  121 ALWEPKSKNASKVANKGKSKS 141
```

RESULT 2

ID   O35473   PRELIMINARY;   PRT;   141 AA.
AC   O35473;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Small unique nuclear receptor co-repressor (Nuclear DNA-binding
DE   protein) (Mus musculus 10 days neonate skin cDNA, RIKEN full-length
DE   enriched library, clone:4732439J08 product:nuclear DNA binding
DE   protein, full insert sequence) (Mus musculus 12 days embryo embryonic
DE   body between diaphragm region and neck cDNA, RIKEN full-length
DE   enriched library, clone:9430096A21 product:nuclear DNA binding
DE   protein, full insert sequence).
GN   Name=Cid; Synonyms=SUN-COR;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CZECH II;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Lazar M.A.;
RT   "Cloning and characterization of a corepressor and potential component
RT   of the nuclear hormone receptor repression complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:14400-14405 (1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CZECH II;
RA   Zamir I., Dawson J., Lavinsky R.M., Glass C.K., Rosenfeld M.G.,
RA   Lazar M.A.;

RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CZECH II;
RC   TISSUE=Mammary tumor metastasized to lung. Tumor arose spontaneously;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and Skin;
RA   Carninci P., Hayashizaki Y.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and Skin;
RX   MEDLINE=21085660; PubMed=11217851;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and Skin;
RX   MEDLINE=22279253; PubMed=10349636;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and Skin;
RA   The FANTOM Consortium;
RA   The RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and Skin;
RX   MEDLINE=20499374; PubMed=11042159;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and Skin;
RX   MEDLINE=20530913; PubMed=11076861;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi Y., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Sumi N., Ishii Y., Nakamura S., Kitsunai T., Taahiro H., Itoh M.,
RA   Yamamoto R., Matsumoto H., Hazama M., Nishine T., Harada A.,
RA   Fujiwake S., Inoue K., Inoue Y., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura T., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [9]
RP   SEQUENCE FROM N.A.

```
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and Skin;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saito R., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF031426; AAC53520.1; -.
DR   EMBL; BC005436; AAH05436.1; -.
DR   EMBL; AK028702; BAC26075.1; -.
DR   EMBL; AK035169; BAC28967.1; -.
DR   MGD; MGI:1927354; Cld.
DR   GO; GO:0017053; C:transcriptional repressor complex; IDA.
DR   GO; GO:0003677; F:DNA binding; IDA.
DR   GO; GO:0016922; F:ligand-dependent nuclear receptor interacto...; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0003714; F:transcription corepressor activity; IDA.
DR   GO; GO:0016481; P:negative regulation of transcription; IDA.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   DNA-binding; Receptor.
SQ   SEQUENCE   141 AA;  15945 MW;  C9AF68033A0FD04E CRC64;

Qy         Match                             Score 654;  DB 2;  Length 141;
Best Local Similarity     90.7%;    Pred. No. 2.3e-50;
Matches   127;    Conservative    8;    Mismatches   5;    Indels    0;    Gaps    0;

Qy      1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          ||||||||||||:|||||||||||||:||||||||||||||||||||||||||||||||||
Db      1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60

Qy     61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120
          ||||||||||||||||||||||||||||||||||||||:||||||||:|||||||||||||
Db     61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKN 120

Qy    121 ALWEPKSKNASKVANKGKSK 140
          ||||||:|: |||||||||||
Db    121 ALWEPKAKSTPKVANKGKSK 140

RESULT 3
ID   Q7TSU0                  PRELIMINARY;     PRT;   141 AA.
AC   Q7TSU0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Nuclear DNA-binding protein C1d.
OS   Cricetulus griseus (Chinese hamster)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Guang L., Masabumi S., Maru Y.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY302220; AAP43113.1; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   DNA-binding.
SQ   SEQUENCE   141 AA;  15945 MW;  0CFB64776AB6566 CRC64;

Query Match                             91.2%;  Score 652;  DB 2;  Length 141;
Best Local Similarity     91.4%;    Pred. No. 3.4e-50;
```

```
Matches   128;    Conservative    4;    Mismatches    8;    Indels    0;    Gaps    0;

Qy      1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          |||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db      1 MAGEEMNEDYPVEIHESLSALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60

Qy     61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120
          |||||||||||||||||||||||||||||||||||||:||||||||:|||||||||||||
Db     61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKN 120

Qy    121 ALWEPKSKNASKVANKGKSK 140
          ||||||:|: |||||||||||
Db    121 ALWEPKQKNTPNVANKGKSK 140

RESULT 4
ID   Q61368                  PRELIMINARY;     PRT;   141 AA.
AC   Q61368;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   C1D protein.
GN   Name=C1d; Synonyms=C1D;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMRI; TISSUE=Ascites tumor;
RX   MEDLINE=98136194; PubMed=9469821;
RA   Nehls P., Kecks T., Greferath R., Spiess E., Glaser T., Rothbarth K.,
RA   Stammer H., Werner D.;
RT   "cDNA cloning, recombinant expression and characterization of
RT   polypeptides with exceptional DNA affinity.";
RL   Nucleic Acids Res. 26:1160-1166(1998).
DR   EMBL; X95591; CAA64844.1; -.
DR   MGD; MGI:1927354; C1d.
DR   GO; GO:0017053; C:transcriptional repressor complex; IDA.
DR   GO; GO:0003677; F:DNA binding; IDA.
DR   GO; GO:0016922; F:ligand-dependent nuclear receptor interacto...; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0003714; F:transcription corepressor activity; IDA.
DR   GO; GO:0016481; P:negative regulation of transcription; IDA.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
SQ   SEQUENCE   141 AA;  16045 MW;  C9AF68033A097161 CRC64;

Query Match                             90.3%;  Score 646;  DB 2;  Length 141;
Best Local Similarity     90.0%;    Pred. No. 1.2e-49;
Matches   126;    Conservative    7;    Mismatches    7;    Indels    0;    Gaps    0;

Qy      1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          |||||||||||||||:|:||:||||||:||||||||||||||||||||||||||||||||
Db      1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60

Qy     61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKN 120
          |||||||||||||||||||||||||||||||||||||:||||||||:|||||||||||||
Db     61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKN 120

Qy    121 ALWEPKSKNASKVANKGKSK 140
          ||||||:||||||||||||||
Db    121 ALWEPKQKNTPNVANKGKSK 140

RESULT 5
ID   Q7L8E1                  PRELIMINARY;     PRT;   111 AA.
AC   Q7L8E1;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
```

DE   BA19501.1 (Novel protein similar to C1D DNA-binding protein).
GN   Name=bA19501.1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lawlor S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL354750; CAD13204.1; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   DNA-binding.
SQ   SEQUENCE   111 AA;   12646 MW;   F563EA631CEC6EE1 CRC64;

Qy   Query Match   76.8%; Score 549; DB 2; Length 111;
     Best Local Similarity   98.2%; Pred. No. 3.9e-41;
     Matches 109; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   31 MLKTMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERI 90
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MLKTMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERI 60

Qy   91 RVYMNRVKEITDKKKAGKLDRGNAASRFVKNALWEPKSKNASKVANKGKSKS 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 RVYMNRVKEITDKKKAGKLDRGNAASRFVKNALWEPKSKNASKVANKGKSKS 111

RESULT 6
ID   Q86F82
AC   Q86F82;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Clone ZZD1007 mRNA sequence.
OS   Schistosoma japonicum (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC   Schistosomatoidea; Schistosomatidae; Schistosoma.
OX   NCBI_TaxID=6182;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22879925; PubMed=12973349;
RA   Hu W., Yan Q., Shen D.K., Liu F., Zhu Z.D., Song H.D., Xu X.R.,
RA   Wang Z.J., Rong Y.P., Zeng L.C., Wu J., Zhang X., Wang J.J., Xu X.N.,
RA   Wang S.Y., Fu G., Zhang X.L., Wang Z.Q., Brindley P.J., McManus D.P.,
RA   Xue C.L., Feng Z., Chen Z., Han Z.G.;
RT   "Evolutionary and biomedical implications of a Schistosoma japonicum
RT   complementary DNA resource.";
RL   Nat. Genet. 35:139-147(2003).
DR   EMBL; AY222982; AAP06005.1; -.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
SQ   SEQUENCE   141 AA;   15966 MW;   DBEFB20FA1894BA0 CRC64;

Qy   Query Match   26.1%; Score 186.5; DB 2; Length 141;
     Best Local Similarity   29.1%; Pred. No. 9.6e-09;
     Matches 39; Conservative   33; Mismatches   59; Indels   3; Gaps   2;

Qy   1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60
          :: || :::|: ||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MGDSSIFDEIPKEISSQLVSFSEATDDVEQLVNKISSFSNNS-SNEVSGLDTVKSELSLC 59

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGK--LDRGNASRFV 118
          |||||||||||||||||||||||||||||||||||||||||::||||| |||||::||
Db   60 YAFNALFMYLRCNGVETQSHPIMQELDRVMNALKRCRSLVEREVSARLTLDKEATTRFV 119

Qy   119 KNALWEPKSKNASK 132
          ||||:||||:||||
Db   120 KHALWKSAHTKTKTK 133

RESULT 7
ID   Q6BJZ2   PRELIMINARY;   PRT;   204 AA.
AC   Q6BJZ2;
DT   01-OCT-2004 (TrEMBLrel. 28, Created)
DT   01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to CA05331IPF19872 Candida albicans IPF19872.
GN   ORFNames=DEHA0F27797g;
OS   Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX   NCBI_TaxID=4959;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RG   GENOLEVURES;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RL   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382138; CAG89900.1; -.
SQ   SEQUENCE   204 AA;   22569 MW;   724A5F99FC34D0C9C CRC64;

Qy   Query Match   22.5%; Score 161; DB 2; Length 204;
     Best Local Similarity   32.7%; Pred. No. 2.7e-06;
     Matches 37; Conservative   27; Mismatches   45; Indels   4; Gaps   2;

Qy   14 IHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLD-PLEQAKVDLVSAYTLNSMFWVYLA 72
          :|: :|| :::|||| :||: ||||||||||||| ||||||||||||||||||||||||
Db   7 VNLFVQSLDNSVDQITEKLEPILKKSLEEKIAASDSQVERIKIYNNYSYVLISILFSYLK 66

Qy   73 TQGVNPKEHPVKQELERIRVYMNRVKEITDK---KKAGKLDRGAASRFVKNAL 122
          |||:|||:|||:|||||||:|||||||||||| ||||||||||||||||||
Db   67 TLGINTDQHPIMKELTRIKLIYMKRYKELEAKLASKDTSKEDAEARTFIQNTL 119

RESULT 8
ID   Q93VA9   PRELIMINARY;   PRT;   217 AA.
AC   Q93VA9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   AT5g25080/T11H3_90 (Hypothetical protein).
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
RA   Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,

RA   Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA   Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA   Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA   Ecker J.R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22088475; PubMed=12093376;
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 3:RESEARCH0029-RESEARCH0029(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA   Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T.,
RA   Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA   Yu G., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY050469; AAK91482.1; -.
DR   EMBL; AY087408; AAM64957.1; -.
DR   EMBL; AF378870; AAK55673.1; -.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   217 AA; 23748 MW; DB44F167BD773DB4 CRC64;

Query Match   21.0%; Score 150.5; DB 2; Length 217;
Best Local Similarity   27.6%; Pred. No. 2.5e-05;
Matches   37; Conservative   35; Mismatches   47; Indels   15; Gaps   2;

Qy   22 ENSIGAVDE------------MLKTMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFVVYL 71
                 :|| | || |:||::|::   : | | |::: ||::|: |
Db   18 ESAIEAVNQTLAYLKELKPQLEQMLTLAEPEVLAAMQPLQRAKTMILLAEATTLYELRL 77

Qy   72 ATQGVNPKEHPVKQELERIRVYMNRVKEITDKKK-----AGKLDRGAASRFVKNALWEPK 126
                | : | |:|:|| :| | |::|:|||||| |||||| ||||||||||
Db   78 RCTGVDPDDHRVKSEIERINVYREKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSLPDLT 137

Qy   127 SKNASKVANKGKSK 140
               : || ||| |||
Db   138 STQKQSIRDLSKGE 151

RESULT 9

ID   Q7QIZ7   PRELIMINARY;   PRT;   169 AA.
AC   Q7QIZ7;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   AgCP3387 (Fragment).
GN   Name=agCG52216; ORFNames=ENSANGG00000016976;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC   EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC   preliminary data.
CC   EMBL; AAAB01008807; EAA04511.1; -.
FT   NON_TER   1   1
SQ   SEQUENCE   169 AA; 19590 MW; 69C2764F057A93C2 CRC64;

Query Match   19.5%; Score 139.5; DB 2; Length 169;
Best Local Similarity   30.7%; Pred. No. 0.00018;
Matches   35; Conservative   28; Mismatches   48; Indels   3; Gaps   3;

Qy   22 ENSIGAVDEMLKTMMSVSRN-ELLQKLDPLEQAKVDLVSAYTLNSMFVVYLLATQG-VNPK 79
                 |: |: | | |:|| |:: : | | | |::|: :|::|: |:| | |
Db   37 ENLMQAIDQIQVTLANVNREYEKYPNFTAEKKVRFDTLCAFCVNSLFWMHEKMLGRPNTV 96

Qy   80 EHPVKQELERIRVYMNRVKEITDK-KKAGKLDRGAASRFVKNALWEPKSKNASK 132
               : :| :| :| |:| : || : | |: | | |||| |::
Db   97 MDDIKADLDRVREAMKRLQTIHDNLTKPRLDQPAAQRFVRAGLYDANNAESDK 150

RESULT 10

O74469

ID   O74469   PRELIMINARY;   PRT;   133 AA.
AC   O74469;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SPCC1739.07 protein.
GN   Name=SPCC1739.07;
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schafer M., Muller-Auer S.,
RA   Gabel C., Fuchs M., Dusterhoft A., Fritzc C., Holzer E., Moestl D.,
RA   Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
RA   Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
DR   EMBL; AL031540; CAA20781.1; -.
DR   PIR; T41115; T41115.
DR   GeneDB_SPombe; SPCC1739.07; -.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
SQ   SEQUENCE   133 AA; 15068 MW; DD33501S0D8D60F1 CRC64;

Query Match   18.7%; Score 133.5; DB 2; Length 133;
Best Local Similarity   27.8%; Pred. No. 0.00047;
Matches   37; Conservative   25; Mismatches   54; Indels   17; Gaps   4;

RESULT 11

```
Qy   16  EYLSAFE---NSIGAVDEMLKTMMSV-SRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYL  71
             ||||    |::| ::|   |     ||:|||: :  | ::|: ::|||:|||:|||
Db    4  EYSELFERLNKQLDNVEDVLKPLKDAESIFELAEGKSELEQAKLYITMSYAINSTLYSFY  63

Qy   72  ATQGVNPKEHPVKQELERIRVYMNRV----KEITDKKAGKLDRGAASRFVKNALWEPKS  127
             ||       ::  |::|          ||| ::   ||  ::::|
Db   64  KLNGIDASERPVMQELQRVKNYISKIQQAEKNVNPKTEAVN------TSNAISSSS     114

Qy  128  KNASKVANKGKSK  140
              |||     ::
Db  115  SNRPKVAKDAATR  127
```

ID   Q6C2H9   PRELIMINARY;   PRT;   194 AA.
AC   Q6C2H9;
DT   01-OCT-2004 (TrEMBLrel. 28, Created)
DT   01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to DEHA0F27797g Debaryomyces hansenii.
GN   ORFNames=YALI0F07755g;
OS   Yarrowia lipolytica (Candida lipolytica).
OC   Eukaryota; Fungi; Ascomycota; Saccharomyceta; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=4952;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RG   GENOLEVURES;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuvglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Miller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RG   GENOLEVURES;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382132; CAG77940.1; -.
SQ   SEQUENCE   194 AA;   21520 MW; D67D79F8068EB3B CRC64;

Query Match   18.6%; Score 133; DB 2; Length 194;
Best Local Similarity   24.8%;
Matches   38; Conservative   31; Mismatches   58; Indels   26; Gaps   4;

RESULT 12

```
Qy   13  EIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLA  72
Db    4  QIEDVLELSHNLQDVTSELSQQIRKMIDFRGVAQ-LPPLEQAQFYSKLAYVTNSAMFAFIL 62

Qy   73  TQGVNPKEHPVKQELERIRVYMNRVKEITDK---------KKAGKLDRGNAASRFV----- 118
Db   63  ASGDPKTHPIMKQLDRVKTYMGRVAHAEGKPGPARKDERNTKVDVPAAKRIIFAQTERA  122

Qy  119  --KNALWEPKS                                                   
Db  123  ISADSIKEPETTSDAAEAFLKDVTKSASKKDKK                            155

Qy       KNASKVANKGKSK  140
Db       KNASKVANKGKSK  140
```

RESULT 13

```
Qy   18  LSAFENSIGA---VDEMLKTMMSVSRN--ELLQKLDPLEQAKVDLVSAYTLNSMFWVYLA  72
Db   18  VSAAEDTLAAAESVGDHLAEMVAAAGEDPDAIAELPPLRRARAFLAMAKAAASLFAVRLR  77

Qy   73  TQGVNPKEHPVKQELERIRVYMNRVKEITDKKA-----GKLDRGAASRFVKNAL--WEP  125
Db   78  CSGVDDEHPIKKFERLSIWEEKLNRFEDWDKAPLRPTTVNTQAAARFIGHSLPHLTT   137

Qy  126  KSKNASKVANKGKSKS  141
Db  138  DQKRSMQAISRGEGGS  153
```

ID   Q6XX09   PRELIMINARY;   PRT;   207 AA.
AC   Q6XX09;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein OSJNBa0091D16.5.
GN   Name=OSJNBa0091D16.5;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Katayose Y.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005731; BAC99871.1; -.
SQ   SEQUENCE   207 AA;   22203 MW; 0B0161AD8190E0071 CRC64;

Query Match   18.5%; Score 132; DB 2; Length 207;
Best Local Similarity   27.9%;
Matches   38; Conservative   33; Mismatches   53; Indels   12; Gaps   4;

ID   BAC99871   PRELIMINARY;   PRT;   207 AA.
AC   BAC99871;
DT   02-MAR-2004 (TrEMBLrel. 27, Created)
DT   02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Katayose Y.;
RT   "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
RT   clone:OSJNBa0091D16.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005731; BAC99871.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   207 AA;   22203 MW; 0B0161AD8190E0071 CRC64;

Query Match   18.5%; Score 132; DB 2; Length 207;
Best Local Similarity   27.9%;
Matches   38; Conservative   33; Mismatches   53; Indels   12; Gaps   4;
```
Qy   18  LSAFENSIGA---VDEMLKTMMSVSRN--ELLQKLDPLEQAKVDLVSAYTLNSMFWVYLA  72
Db   18  VSAAEDTLAAAESVGDHLAEMVAAAGEDPDAIAELPPLRRARAFLAMAKAAASLFAVRLR  77

Qy   73  TQGVNPKEHPVKQELERIRVYMNRVKEITDKKA-----GKLDRGAASRFVKNAL--WEP  125
```

```
Db

Qy      78 CSGVDPDEHPIKKFPERLSLWEEKLNRFEDWDKAPLRPTTTVNTQAAARPIGHSLPHLTT 137
            |:|:| |::|: | ::|:|
Db     138 DQKRSMQAISRGEGGS 153

RESULT 14
ID   Q9N3J4
QN   Q9N3J4;
AC   Q9N3J4;                    PRELIMINARY;      PRT;    133 AA.
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein Y51H7C.7.
GN   ORFNames=Y51H7C.7;
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Wilson R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Bradshaw-Cordum H., DuBuque T.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.H.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
RN   [12]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
RN   [13]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [14]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Wilson R.;
RL   Submitted (NOV-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC024805; AAK39333.1; -.
DR   WormPep; Y51H7C.7; CB22363.
KW   Hypothetical protein.
SQ   SEQUENCE 133 AA; 14852 MW; 1F92FB2387391CBF CRC64;

Query Match             18.0%;  Score 128.5; DB 2;  Length 133;
Best Local Similarity   29.3%;  Pred. No. 0.0013;
Matches   36;  Conservative   29; Mismatches   47;  Indels   11;  Gaps    3;

Qy      16 EYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLLVSAYTLNSMPWVYLATQG 75
            || :|:|| || :||| : |||
Db      20 ELITKLEDAVEEVDVGVEKHFERSAHEM------ALVDTMSMFLMDSLMWAVAQATKG 70

Qy      76 VN-PKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVKNALWE-PKSKNASKV 133
            || |:|| :| :|::: || |:::|| |::|| | ||||: |:||||||| ||||||
Db      71 GGADKNDDLLIDLARTKRMTADMKEINLRQDAPRINKQAAANFVRNALWEQPEQGESSKK 130

Qy     134 ANK 136
            |:|
Db     131 AAK 133

RESULT 15
ID   Q9VXL4
QN   Q9VXL4;
AC   Q9VXL4;                    PRELIMINARY;      PRT;    159 AA.
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG8928-PA.
GN   ORFNames=CG8928;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
```

```
Matches  40;  Conservative  18;  Mismatches  58;  Indels  15;  Gaps  2;

Qy    1 MAGEEINEDYPVEIHHEYLSAFENSIGAVDEMLKTMMSVSRNELLQ---------KL  47
          |||| |    ||| | | | |  |      |  |  :  ||:        :    KL
Db    1 MAQENQAVDNGLPCNAYLDTSLREDENMQHILKTF--YSSIELLEADTEKAlALQAERTL  58

Qy   48 DPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAG 107
          | ||  | | :  |||| | | ||: || :| | | :|| :|| :| :  : | |||| |
Db   59 NTNEQIKLDSYLVYLNSTLFFIYLKLQGEDASNHAVMHIDLRRTRDJLJARDKKINDALAAP 118

Qy  108 KLDRGAASRFV 118
          :||  ||:||:
Db  119 RLDMPAAKRFI 129
```

Search completed: January 4, 2005, 13:44:39
Job time: 76.5 secs

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Nelson D.R.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE003500; AAF48545.1; -.
DR IntAct; Q9VXL4; -.
DR FlyBase; FBgn0030711; CG8928.
SQ SEQUENCE 159 AA; 18138 MW; 38CB9E3O2A75BFD6 CRC64;

Query Match 17.7%; Score 126.5; DB 2; Length 159;
Best Local Similarity 30.5%; Pred. No. 0.0024;

QM protein - nucleic search, using frame_plus_p2n model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: January 5, 2005, 12:51:43

2055.201 Million cell updates/sec

Title: US-09-701-618A-2
Perfect score: 715
Sequence: 1 MAGEEINEDYPVEIHEYLSA..........LWEPKSKNASKVANKGKSKS 141

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09701618/runat_05012005_085839_24447/app_query.fasta_1.654
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09701618 @CGN_1_1.6628 @runat_05012005_085839_24447 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7 -YGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

Database :   EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Length	DB	ID	Description
1	715	100.0	497	6	CA867549	CA867549 ir29g10.y
2	715	100.0	498	6	CB160277	CB160277 K-EST0219
3	715	100.0	521	5	BQ636280	BQ636280 hd06g11.y
4	715	100.0	521	6	CD101786	CD101786 AGENCOURT
5	715	100.0	524	7	CK430831	CK430831 oj54d12.y
5	715	100.0	529	7	CF455458	CF455458 AGENCOURT
6	715	100.0	563	4	BG496371	BG496371 602541062
7	715	100.0	653	4	BG677242	BG677242 602623944
8	715	100.0	675	7	CK001081	CK001081 AGENCOURT

10	715	100.0	738	5	BU561215	BU561215 AGENCOURT
11	715	100.0	756	4	BG697814	BG697814 602661015
12	715	100.0	771	5	BU928625	BU928625 AGENCOURT
13	715	100.0	789	6	CB998397	CB998397 AGENCOURT
14	715	100.0	813	6	BU568056	BU568056 AGENCOURT
15	715	100.0	838	5	BU568457	BU568457 AGENCOURT
16	715	100.0	872	6	BE886831	BE886831 601507804
17	715	100.0	963	6	CD515525	CD515525 AGENCOURT
18	715	100.0	979	5	BQ219896	BQ219896 AGENCOURT
19	715	99.3	469	7	AA424253	AA424253 zv1le12.r
20	715	99.3	555	4	BG283692	BG283692 602407460
21	715	99.3	647	6	CD688633	CD688633 EST155 h
22	710	99.3	656	5	CB138723	CB138723 K-EST0191
23	710	99.3	788	1	AV763046	AV763046 AGENCOURT
24	710	99.3	982	5	BQ434640	BQ434640 AGENCOURT
25	710	99.3	988	4	BM923929	BM923929 AGENCOURT
26	710	99.3	1014	4	BM451300	BM451300 AGENCOURT
27	708	99.1	613	4	BG533158	BG533158 602580684
28	707	98.9	773	6	CB231101	CB231101 AGENCOURT
29	707	98.9	781	6	CB230406	CB230406 AGENCOURT
30	707	98.9	794	6	CB309384	CB309384 AGENCOURT
31	707	98.9	805	6	CB312445	CB312445 AGENCOURT
32	707	98.9	915	7	CN645648	CN645648 ILLUMIGEN
33	706	98.9	470	4	BM129833	BM129833 1f23g06.y
34	706	98.7	602	1	AV685739	AV685739 AGENCOURT
35	705	98.7	654	1	AV686813	AV686813 AGENCOURT
36	705	98.7	542	2	BF667890	BF667890 60122126
37	704	98.5	807	5	BU853132	BU853132 AGENCOURT
38	702	98.2	577	4	BG502972	BG502972 602550641
39	702	98.2	779	7	CF596623	CF596623 AGENCOURT
40	702	98.2	787	6	CD000617	CD000617 AGENCOURT
41	700.5	98.2	805	6	BF218466	BF218466 60188183.9
42	700	97.9	784	2	CN342946	CN342946 170005322
43	699	97.8	464	7	AI935349	AI935349 wo82d07.x
c 44	698	97.6	609	1	BG501950	BG501950 602549042
45	698	97.6	667	4	BG613266	BG613266 602641127

ALIGNMENTS

RESULT 1
CA867549
LOCUS   CA867549   497 bp   mRNA   linear   EST 20-DEC-2002
DEFINITION ir29g10.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6546812 5'
            similar to TR:Q13901 Q13901 CID DNA-BINDING PROTEIN. ;, mRNA
            sequence.
ACCESSION   CA867549
VERSION   CA867549.1 GI:27319098
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 497)
AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Bistalin,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
            Endocrine Pancreas Consortium
TITLE   Endocrine Pancreas Consortium
JOURNAL   Unpublished (2000)
COMMENT   Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue

RESULT 2

CB160277

LOCUS   CB160277   498 bp   mRNA   linear   EST 30-JAN-2003
DEFINITION K-EST0219954 L18POOLin1 Homo sapiens cDNA clone L18POOLin1-52-H02
            5', mRNA sequence.
ACCESSION   CB160277
VERSION   CB160277.1 GI:28146403
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 498)
AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE   21C Frontier Korean EST Project 2001
JOURNAL   Unpublished (2002)
COMMENT   Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            /clone_lib="L18POOLin1"
            /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
            Site_2: NotI; The library was contributed by the Soares
            laboratory and it was constructed as described by Bonaldo,
            M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
            6(9): 791-806. RNA was prepared from harvested cell
            culture."

FEATURES   Location/Qualifiers
     source   1..498
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /clone="L18POOLin1-52-H02"
                     /cell_line="SNU-354+Cho-CK-Choi-CK+HLK-3"
                     /lab_host="top10F'"
                     /clone_lib="L18POOLin1"
                     Plate: 52 row: H column: 02
                     High quality sequence stop: 498.

ORIGIN

Alignment Scores:
Pred. No.:
Score: 715.00
Percent Similarity: 100.00%   Length: 498
Best Local Similarity: 100.00%   Matches: 141
Query Match: 100.00%   Conservative: 0
DB: Mismatches: 0
                                        Indels: 0
                        1.6e-80   Gaps: 0

US-09-701-618A-2 (1-141) x CB160277 (1-498)

```
Qy    1 MetAlaGlyGluGluIleIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGGTAGAAATTCACGAGTAATTTGTCAGCG 95

Qy   21 PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96 TTTGAGAATTCCATTGGTGCTGTGGATGAGAGATGCTGAGACAGTGTCTGTTTCAGA 155

Qy   41 AsnGluLeuGlnLysLeuAspProLeuGlnAlaLysValAspLeuValSerAla 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  156 AATGAGTTGTTGCAGAAGTTGGATCCACTGCAAGCAAAGGTGGATTTGGTTTCTGCA 215

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  216 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 275

Qy   81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  276 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 335

Qy  101 ThrAspLysLysLysAlaGlyLysLysLeuAspArgGlyValAlaAlaSerArgPheValLysAsn 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  336 ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAT 395

Qy  121 AlaLeuTrpGluProLysSerLysValAlaAsnLysGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  396 GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAGGTTGCCAATAAAGGAAAAAGTAA 455
```

(hinoue@im.wustl.edu)
Seq primer: -400P from Gibco
High quality sequence stop: 487.

FEATURES   Location/Qualifiers
     source   1..497
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /clone="IMAGE:6546812"
                     /db_xref="taxon:9606"
                     /tissue_type="Purified pancreatic islet"
                     /clone_lib="HR85 islet"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."

ORIGIN

Alignment Scores:
Pred. No.:
Score: 715.00
Percent Similarity: 100.00%   Length: 497
Best Local Similarity: 100.00%   Matches: 141
Query Match: 100.00%   Conservative: 0
DB: Mismatches: 0
                                        Indels: 0
                        1.59e-80   Gaps: 0

US-09-701-618A-2 (1-141) x CA867549 (1-497)

```
Qy    1 MetAlaGlyGluGluIleIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   48 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGGTAGAAATTCACGAGTAATTTGTCAGCG 107

Qy   21 PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  108 TTTGAGAATTCCATTGGTGCTGTGGATGAGAGATGCTGAGACAGTGTCTGTTTCAGA 167

Qy   41 AsnGluLeuGlnLysLeuAspProLeuGlnAlaLysValAspLeuValSerAla 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  168 AATGAGTTGTTGCAGAAGTTGGATCCACTGCAAGCAAAGGTGGATTTGGTTTCTGCA 227

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  228 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 287

Qy   81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  288 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 347

Qy  101 ThrAspLysLysLysAlaGlyLysLysLeuAspArgGlyValAlaAlaSerArgPheValLysAsn 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  348 ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAT 407

Qy  121 AlaLeuTrpGluProLysSerLysValAlaAsnLysGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  408 GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAGGTTGCCAATAAAGGAAAAAGTAA 467

Qy  141 Ser 141
        |||
Db  468 AGT 470
```

```
Qy     141 Ser 141
           |||
Db     456 AGT 458
```

RESULT 3

```
Db   BQ636280
     BQ636280
LOCUS       BQ636280     521 bp    mRNA    linear    EST 15-JUL-2002
DEFINITION  hd06g11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
            Homo sapiens cDNA clone hd06g11 5', mRNA sequence.
ACCESSION   BQ636280
VERSION     BQ636280.1  GI:21760739
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
            Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A.,
  TITLE     Expressed sequence tag analysis of human retina for the NEIBank
            Project: Retbindin, an abundant, novel retinal cDNA and alternative
            splicing of other retina-preferred gene transcripts
  JOURNAL   Mol. Vis. 8 (4), 196-204 (2002)
  MEDLINE   22103461
   PUBMED   12107411
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 06  row: g  column: 11
            Seq primer: M13Rp1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1..521
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="hd06g11"
                     /tissue_type="Retina"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /clone_lib="Human Retina cDNA (Un-normalized,
                     unamplified): hd/he"
                     /note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
                     was dissected from two 80 year old donors with no observed
                     eye disease. 100ug of total RNA was used for library
                     construction. A directionally cloned cDNA library in the
                     pSPORT1 vector (Life Technologies) was constructed at
                     Bioserve Biotechnology (Laurel MD) essentially following
                     the protocols of the SuperScript Plasmid System full
                     details of which are contained in the manufacturer's
                     Instruction manual (http://www.lifetech.com/). First
                     strand synthesis was carried out using a Not I
                     primer-adapter
                     [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCC(T)15-3']. EST analysis
                     was performed on the unamplified library at the NIH
                     Intramural Sequencing Center (NISC)."
ORIGIN
```

Alignment Scores:

```
Pred. No.:                                Length:     521
Score:              1.7e-80               Matches:    141
Percent Similarity: 715.00                Conservative: 0
Best Local Similarity: 100.00%            Mismatches:   0
Query Match:        100.00%               Indels:       0
DB:                 100.00%               Gaps:         0
                    5
```

US-09-701-618A-2 (1-141) x BQ636280 (1-521)

```
Qy   1 MetAlaGlyGluGlyIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
```

```
Db  59 ATGGCAGGTGAAGGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 118
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
Db 119 TTTGAGAATTCCATTCGGTGCTGTGGATGAGATGGAAGACCATGAGTCTCGTTTCAGA 178

Qy  41 AsnGluLeuLeuGlnLysValAspProLeuGluGluGlnAlaLysValAlaAspLeuValSerAla 60
Db 179 AATGAGTTGTTGCAGAAAGTTGGATCCACTTGAACAAGCAAAGAGTGGATTGGTTCTCGCA 238

Qy  61 TyrThrLeuAsnSerMetPheTrpValTyrLeuValLaThrHisGlnGlyValAlaProGlyGlu 80
Db 239 TACACATTAAATTCAATGTTTTGGGTTTATTTGGTT||||||||CCAAGGAGGTTAATCCTAAGGAA 298

Qy  81 HisProValLysGlyGlnLeuGluArgValTyrMetAsnArgValLysGluIle 100
Db 299 CATCCAGTAAAACAGGAATTCGAAAGAGTATATCGAACACGAGTCAAGGAAATA 358

Qy 101 ThrAspGlyLysGlyAlaGlyLysLysLeuAspArgGlyValAlaAlaSerArgPheValLysAsn 120
Db 359 ACAGACAAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGAATTTGTAAAAAT 418

Qy 121 AlaLeuTrpGluProLysSerLysLysGlnAlaSerArgLysGlyGlyLysSerLys 140
Db 419 GCCCTCTGGGGAACCAAAATCGAAAAAGTTGCCAATAAAGGAAAAAGTAAA 478

Qy 141 Ser 141
        |||
Db 479 AGT 481
```

RESULT 4

```
Db   CD101786
     CD101786
LOCUS       CD101786     521 bp    mRNA    linear    EST 15-MAY-2003
DEFINITION  AGENCOURT 13980951 NIH_MGC_187 Homo sapiens cDNA clone
            IMAGE:30373459 5', mRNA sequence.
ACCESSION   CD101786
VERSION     CD101786.1  GI:30754960
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM167  row: e  column: 20
            High quality sequence stop: 521.
FEATURES             Location/Qualifiers
     source          1..521
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30373459"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_187"
                     /note="Organ: Blood vessels - aorta, basilar and artery;
                     Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2:
                     SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in
                     cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
```

```
Qy 141 Ser 141
        |||
Db 479 AGT 481
```

Column 1

ORIGIN

(range 0.5-4.0 kb). 14/15 colonies contained inserts by
PCR. This library was enriched for full-length clones and
was constructed by Clontech Laboratories (Palo Alto, CA).
Note: this is a NIH_MGC library."

US-09-701-618A-2 (1-141) x CD101786 (1-521)

Alignment Scores:

Pred. No.:			
Score:	715.00	Length:	521
Percent Similarity:	100.00%	Matches:	141
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	100.00%	Mismatches:	0
DB:		Indels:	0
		Gaps:	0

```
Qy    1 MetAlaGlyGluGluIleLeuAsnGluAspTyrProValGluIleIleHisGluTyrLeuSerAla  20
Db   54 ATGGCAGGTGAAGAAATTAATGAAGACTACCCAGTAGAAATTCACGAGTATTTGTCAGCG        113
Qy   21 PheGluAsnSerIleGlyAlaValAlaAspThrMetLeuLysThrMetSerValSerArg        40
Db  114 TTTGAGAATTCCATTGGTGCTGTGGCAGACACCATGAGCGTCTCGTTTCTAGA              173
Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnGlnAlaLysValSerAla                 60
Db  174 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGTTTCTGCA             233
Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuValAlaThrGlnGlyValAsnProLysGlu     80
Db  234 TACACATTAAATTCAAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGA       293
Qy   81 HisProValLysGlnLeuGluLeuGluArgIleTyrMetAsnArgValLysGluIle          100
Db  294 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTAIATATGAACAGAGTCAAGGAAATA       353
Qy  101 ThrAspLysGlyLysLeuLysLeuAspArgGlyValAlaAlaSerArgPheValLysSerAsn    120
Db  354 ACAGACAAGAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAGT         413
Qy  121 AlaLeuTrpGluProLysSerIleSerValAlaAlaSerLysAsnLysGlyLysSerLys       140
Db  414 GCCCCTCTGGGAACCAAAAATCGAAAAAGTTGCCAATAAAGGAAAAAAGTAAA             473
Qy  141 Ser                                                               141
Db  474 AGT                                                               476
```

RESULT 5

LOCUS   CK430831   524 bp   mRNA   linear   EST 06-JAN-2004

DEFINITION   clone oj54d12 5', unamplified: oj Homo sapiens cDNA clone oj54d12 5', mRNA sequence.

ACCESSION   CK430831

VERSION   CK430831.1   GI:40679320

KEYWORDS   EST.

SOURCE   Homo sapiens (human)

ORGANISM   Homo sapiens
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 524)

AUTHORS   Dickinson,D., Laurie,G. and Wistow,G.

TITLE   Expressed sequence tag analysis of human lacrimal gland

JOURNAL   Unpublished (2002)

COMMENT   Contact: Wistow G
     National Eye Institute
     Section on Molecular Structure and Function
     6/331, NIH, Bethesda, MD 20892-2740, USA
     Tel: 301 402 3452
     Fax: 301 496 0078
     Email: graeme@helix.nih.gov
     Plate: 54 row: d column: 12
     Seq primer: M13R1 reverse primer (ABI).

Column 2

FEATURES   Location/Qualifiers

source   1..524
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /clone="oj54d12"
     /db_xref="taxon:9606"
     /tissue_type="lacrimal gland"
     /dev_stage="Adult"
     /lab_host="EMDH10B"
     /clone_lib="Human lacrimal gland, unamplified; oj"
     /note="Organ: Eye; Vector: pCMVSport6; RNA was extracted
     from 2 human lacrimal glands. A directionally cloned cDNA
     library in the pCMVSPORT6 vector(Life Technologies) was
     constructed at Bioserve Biotechnology (Laurel MD)
     essentially following the protocols of the SuperScript
     Plasmid System full details of which are contained in the
     manufacturer's Instruction manual
     (http://www.lifetech.com/). First strand synthesis was
     carried out using a Not I primer-adapter
     [5'-pGACTAGTTCTAGAtCGCGAGCGGCGCGCGCCC(T)15-3'. EST analysis
     was performed on the unamplified library at the NIH
     Intramural Sequencing Center (NISC)."

ORIGIN

US-09-701-618A-2 (1-141) x CK430831 (1-524)

Alignment Scores:

Pred. No.:			
Score:	715.00	Length:	524
Percent Similarity:	100.00%	Matches:	141
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	100.00%	Mismatches:	0
DB:		Indels:	0
		Gaps:	0

```
Qy    1 MetAlaGlyGluGluIleLeuAsnGluAspTyrProValGluIleIleHisGluTyrLeuSerAla  20
Db   73 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG        132
Qy   21 PheGluAsnSerIleGlyAlaValAlaAspThrMetLeuLysThrMetSerValSerArg        40
Db  133 TTTGAGAATTCCATTGGTGCTGTGGCAGACACCATGAGCGTCTCGTTTCTAGA              192
Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnGlnAlaLysValSerAla                 60
Db  193 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGTTTCTGCA             252
Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuValAlaThrGlnGlyValAsnProLysGlu     80
Db  253 TACACATTAAATTCAAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGA       312
Qy   81 HisProValLysGlnLeuGluLeuGluArgIleTyrMetAsnArgValLysGluIle          100
Db  313 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTA                               372
Qy  101 ThrAspLysGlyLysLeuLysLeuAspArgGlyValAlaAlaSerArgPheValLysSerAsn    120
Db  373 ACAGACAAGAAAGGCTGGCAAGCTGGACAGAGG                                  432
Qy  121 AlaLeuTrpGluProLysSerIleSerValAlaAlaSerLysAsnLysGlyLysSerLys       140
Db  433 GCCCCTCTGGGAACCAAAAATCGAAAAAGT                                     492
Qy  141 Ser                                                               141
Db  493 AGT                                                               495
```

RESULT 6

LOCUS   CF455458   529 bp   mRNA   linear   EST 04-SEP-2003

DEFINITION   AGENCOURT 15332424 Lupski_anterior_horn Homo sapiens cDNA clone IMAGE:30516205 5', mRNA sequence.

ACCESSION   CF455458

VERSION   CF455458.1   GI:34455114

KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 529)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM582 row: 1 column: 14
            High quality sequence stop: 529.
FEATURES   Location/Qualifiers
     source   1..529
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3016205"
                     /tissue_type="Peripheral Nervous system"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="Lupski anterior horn"
                     /note="Vector: pCMV-SPORT6.1; Site_1: EcoRV (destroyed);
                     Site_2: NotI; Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 2.1 kb. Library was constructed by Invitrogen
                     and donated by J. Lupski, M.D./Ph.D. (Baylor College of
                     Medicine)."
ORIGIN

US-09-701-618A-2 (1-141) x CF455458 (1-529)

Alignment Scores:
Pred. No.:
Score: 1.73e-80   Length: 529
Score: 715.00   Matches: 141
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 7   Gaps: 0

Qy   1 MetAlaGlyGluGluIleAsnGluAspTyrProValGlyIleHisGluTyrLeuSerAla   20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 102

Qy   21 PheGluAsnSerIleGlyAlaValAlaSpGluMetLeuLysThrMetSerValSerArg   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 103 TTTGAGAATTCAATGGGTGCTGTGGTAGATGAGATGCTGAAGACCATGAGTGTCTGTTCTAGA 162

Qy   41 AsnGluLeuLeuGlnLysGlyLeuAspProLeuGluGlnGlnAlaLysValIleAspGluAla   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 163 AATGAGTTGTTGCAGAAGGTTGGATCCACTTGAACAAGCAAAAGTGGAATTTGGTTTCTGCA 222

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu   80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 223 TACACATTAAATCAATGTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 282

Qy   81 HisProValLysGlnGluLeuGlnArgIleArgValTyrMetAsnArgValLysGluIle 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 283 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGACTATATAATGAACAGAGTCAAGGAAATA 342

Qy 101 ThrAspLysLysGlyAlaGlyLysLeuAspArgGlyTyrAlaAlaSerArgPheValLysAsn 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 343 ACAGACAAGAAAAAGGCTGGCAAGCTTGACAGAGGGTGCAGCTTCAAGATTTGTAAAAAAT 402

REFERENCE (right column)

Qy 121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 403 GCCCTCTGGGAACCAAAATCCAAAATGCATCAAAAAGTTGCCAATAAAAGGAAAAGTAAA 462

Qy 141 Ser 141
          |||
Db 463 AGT 465

RESULT 7
BG496371
BG496371
LOCUS   BG496371   563 bp   mRNA   linear   EST 27-MAR-2001
DEFINITION 602541062F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4672262 5',
            mRNA sequence.
ACCESSION   BG496371
VERSION   BG496371.1 GI:13457888
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 563)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1488 row: k column: 15
            High quality sequence stop: 536.
FEATURES   Location/Qualifiers
     source   1..563
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4672262"
                     /tissue_type="mucoepidermoid carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_59"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
ORIGIN

US-09-701-618A-2 (1-141) x BG496371 (1-563)

Alignment Scores:
Pred. No.:
Score: 1.88e-80   Length: 563
Score: 715.00   Matches: 141
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 4   Gaps: 0

Qy   1 MetAlaGlyGluGluIleAsnGluAspTyrProValGlyIleHisGluTyrLeuSerAla   20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75 ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 134

```
Pred. No.:                          Length:        653
Score:                              Matches:       141

Alignment Scores:   2.28e-80
                    715.00

ORIGIN
```

RESULT 8

```
BG677242
BG677242
LOCUS       BG677242      653 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  602623944F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4748963 5',
            mRNA sequence.
ACCESSION   BG677242
VERSION     BG677242.1  GI:13908639
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10601 row: g column: 12
            High quality sequence stop: 649.
FEATURES            Location/Qualifiers
     source         1..653
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4748963"
                    /tissue_type="squamous cell carcinoma"
                    /lab_host="DH10B (T1 phage-resistant)"
                    /clone_lib="NCI_CGAP_Skn4"
                    /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                    Average insert size 1.5kb. Library constructed by Life
                    Technologies. Note: this is a NCI_CGAP Library."
```

```
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:     4                                Gaps:           0

US-09-701-618A-2 (1-141) x BG677242 (1-653)
```

```
Qy   21  PheGluAsnSerIleGlyTyrAlaValAlaAspGluMetLeuLysThrMetSerValSerArg  40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  135  TTTGAGAATTCCATTGGTCGTGTGATGAGAATGCTGAAGACCATGAGTGTCTGTTTCTAGA  194

Qy   41  AsnGluLeuLeuGlnLysLeuLysAspProLeuGluGlnAlaLysValAlaSpLeuValSerAla  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  195  AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGCATTGGTTTCTGCA  254

Qy   61  TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAlaAsnProLysGlu  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  255  TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTGCAAATCCTAAGGAA  314

Qy   81  HisProValLysGlnGlnLeuLeuGluLeuArgIleArgValTyrMetAsnArgValLysGluIle  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  315  CATCCAGTAAAACAGCAGTTGCTGGAGTTAAGGATTCGAGTTTATATGAACAGAGTCAAGGAAATA  374

Qy  101  ThrAspLysLysLysAlaGlyLysLysLeuLeuAspArgGlyAlaAlaSerArgPheValLysAsn  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  375  ACAGACAAGAAAAAGGCTGGCAAGAAGCTTCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAT  434

Qy  121  AlaLeuTrpGluProLysSerLysSerAsnLysLysValLysSerLys  140
         ||||||||||||||||||||||||||||||||||||||||||||
Db  435  GCCCTCTGGGAACCAAAATCGAAAAGTAAGAAAAAGTAA  494

Qy  141  Ser  141
         |||
Db  495  AGT  497
```

RESULT 9

```
CK001081
CK001081
LOCUS       CK001081      675 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT 16390846 NIH_MGC_227 Homo sapiens cDNA clone
            IMAGE:30718850 5', mRNA sequence.
ACCESSION   CK001081
VERSION     CK001081.1  GI:38527115
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Professor Miklos Palkovits
            cDNA Library Preparation: Michael Brownstein / Ted Usdin
            Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM262 row: m column: 03
            High quality sequence stop: 545.
FEATURES            Location/Qualifiers
     source         1..675
                    /organism="Homo sapiens"
```

```
Qy    1  MetAlaGlyGluGluIleAsnGluAspTyrProValGlyIleGluHisGluTyrLeuSerAla  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   36  ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGGAATTGAACATGAATATTTGTCAGCG  95

Qy   21  PheGluAsnSerIleGlyTyrAlaValAlaAspGluMetLeuLysThrMetSerValSerArg  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   96  TTTGAGAATTCCATTGGTCGTGTGATGAGAATGCTGAAGACCATGAGTGTCTGTTTCTAGA  155

Qy   41  AsnGluLeuLeuGlnLysLeuLysAspProLeuGluGlnAlaLysValAlaSpLeuValSerAla  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  156  AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGCATTGGTTTCTGCA  215

Qy   61  TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAlaAsnProLysGlu  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  216  TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTGCAAATCCTAAGGAA  275

Qy   81  HisProValLysGlnGlnLeuLeuGluLeuArgIleArgValTyrMetAsnArgValLysGluIle  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  276  CATCCAGTAAAACAGCAGTTGCTGGAGTTAAGGATTCGAGTTTATATGAACAGAGTCAAGGAAATA  335

Qy  101  ThrAspLysLysLysAlaGlyLysLysLeuLeuAspArgGlyAlaAlaSerArgPheValLysAsn  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  336  ACAGACAAGAAAAAGGCTGGCAAGAAGCTTCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAT  395

Qy  121  AlaLeuTrpGluProLysSerLysSerAsnLysLysValLysSerLys  140
         ||||||||||||||||||||||||||||||||||||||||||||
Db  396  GCCCTCTGGGAACCAAAATCGAAAAGTAAGAAAAAGTAA  455

Qy  141  Ser  141
         |||
Db  456  AGT  458
```

ORIGIN
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30718850"
/tissue_type="Bulk tissue from Human Spinal cord"
/lab_host="DH10B TonA"
/clone_lib="NIH_MGC 227"
/note="Organ: brain/CNS; Vector: pDNR-LIB; Site_1: SfiI
(directional); Site_2: SfiI (directional); Library is
oligo-dT primed and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGG-3'
5'-ATTCTAGAGGCCGAGGCGGCCGACATG-d(T)30N-1N-3. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected for >0.5kb with an
average insert size of 1.3kb Library created in the
laboratory of Jonathan Kuo and Ted Usdin."

US-09-701-618A-2 (1-141) x CK001081 (1-675)

Alignment Scores:
Pred. No.:
Score: 2.38e-80
Percent Similarity: 715.00   Length: 675
Best Local Similarity: 100.00%   Matches: 141
Query Match: 100.00%   Conservative: 0
DB: 7   Mismatches: 0
                                            Indels: 0
                                            Gaps: 0

Qy   1 MetAlaGlyGluGlyIleLeuAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 ATGGCAGGTGAAGGAATTAATGAACGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 125

Qy   21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 126 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGAGTGTCTGTTTCTAGA 185

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnAlaLysValAlaSerAla 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 186 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCGCA 245

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAlaAsnProLysGlu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246 TACACATTAAATTCAAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 305

Qy   81 HisProValLysGlnGlnLeuGluArgIleLeuArgValTyrMetAsnArgValLysGluIle 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 306 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 365

Qy 101 ThrAspLysGlyLysAlaGlyLysLysLeuAspArgArgGlyAlaAlaSerArgPheValLysAsn 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 366 ACAGACAAGAAAAGGCTGGACAGGTGCACGCTTCAAGATTTGTAAAAAT 425

Qy 121 AlaLeuTrpGluProLysGlySerLysSerLys 140
        ||||||||||||||||||||||||||||||||||
Db 426 GCCCTCTGGGAACCAAAAATCGAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAAGTAAA 485

Qy 141 Ser 141
        |||
Db 486 AGT 488

REFERENCE   1 (bases 1 to 738)

AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
           National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   CDNA Clone Sequence Inc.
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: CLONTECH
           cDNA Library Preparation: CLONTECH
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLcM2814 row: o column: 11
           High quality sequence stop: 560.

FEATURES   Location/Qualifiers
     source   1..738
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:6592163"
                    /lab_host="DH10B (T1 phage-resistant)"
                    /clone_lib="NIH_MGC 82"
                    /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                    SfiI (ggccgcccggcc); Site_2: SfiI (ggccattatggcc); 5' adaptor
                    3' adaptors were used in cloning as follows: 5' adaptor
                    sequence: 5'-CACGGCGCCATPATGGCC-3', and 3' adaptor sequence:
                    5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                    C, or G and N = A, C, G, or T). Average insert size
                    1.35 kb (range 0.9-4.0 kb). This library was enriched for full-length
                    inserts by PCR. This library was enriched for full-length
                    clones and was constructed by Clontech Laboratories (Palo
                    Alto, CA)."

ORIGIN

US-09-701-618A-2 (1-141) x BU561215 (1-738)

Alignment Scores:
Pred. No.:
Score: 2.67e-80
Percent Similarity: 715.00   Length: 738
Best Local Similarity: 100.00%   Matches: 141
Query Match: 100.00%   Conservative: 0
DB: 5   Mismatches: 0
                                            Indels: 0
                                            Gaps: 0

Qy   1 MetAlaGlyGluGlyIleLeuAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   46 ATGGCAGGTGAAGGAATTAATGAACGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 105

Qy   21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 106 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGAGTGTCTGTTTCTAGA 165

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnAlaLysValAlaSerAla 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 166 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCGCA 225

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAlaAsnProLysGlu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 226 TACACATTAAATTCAAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 285

Qy   81 HisProValLysGlnGlnLeuGluArgIleLeuArgValTyrMetAsnArgValLysGluIle 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 345

Qy 101 ThrAspLysGlyLysAlaGlyLysLysLeuAspArgArgGlyAlaAlaSerArgPheValLysAsn 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 346 ACAGACAAGAAAAGGCTGGACAGGTGCACGCTTCAAGATTTGTAAAAAT 405

Qy 121 AlaLeuTrpGluProLysGlySerLysSerLys 140
        ||||||||||||||||||||||||||||||||||
Db 406 GCCCTCTGGGAACCAAAAATCGAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAAGTAAA 465

Qy 141 Ser 141

LOCUS   BU561215   738 bp   mRNA   linear   EST 16-SEP-2002
DEFINITION AGENCOURT_10279216 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6592163
            5', mRNA sequence.
ACCESSION   BU561215
VERSION   BU561215.1 GI:22911511
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Db

RESULT 11

BG697814

```
LOCUS       BG697814                756 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602661015F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4804394 5',
            mRNA sequence.
ACCESSION   BG697814
VERSION     BG697814.1  GI:13964449
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 756)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM0701 row: m column: 03
            High quality sequence stop: 683.
FEATURES             Location/Qualifiers
     source          1..756
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /clone="IMAGE:4804394"
                     /clone_lib="NCI_CGAP_Skn3"
                     /db_xref="taxon:9606"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
```

US-09-701-618A-2 (1-141) x BG697814 (1-756)

```
Alignment Scores:
Pred. No.:
Score:                 715.00          Length:        756
Percent Similarity:    100.00%         Matches:       141
Best Local Similarity: 100.00%         Conservative:    0
Query Match:           4               Mismatches:      0
                                       Indels:          0
                                       Gaps:            0
```

ORIGIN

```
Qy    1 MetAlaGlyGluGluIleAsnGluAspProLeuGluGlnAlaLysValAspLeuLeuSerAla 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 ATGGCAGGTGAAGAAATTAATGAAGACTATCCACTTGAACAAGCAAAAGTGGATTTGTTTCTCGCA 129

Qy   21 PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  130 TTTGAGAATTCAATTGGTGCTGTTGCTGATGAAATGCTGAAGACAATGTCTGTTTCTCGA 189

Qy   41 AsnGluLeuLeuGlnLysValAspLeuAlaLysSerValAla 60
        ||||||||||||||||||||||||||||||||||||||||
Db  190 AATGAGTCGTTGCAGAAGTTGGATCCACTTGACAAGCAAAAAGTGGATTTGTTTTCTCGCA 249

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  250 TACACATTAAATTCAATGTTTCGGTTTATTCGGCAACCAAGGAGTTAATCCTAAGGAA 309

Qy   81 HisProValLysGlnLeuLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  310 CATCCAGTAAACAGGAATTCAGAGTATAATATGAACAGAGTCAAGGAAATA 369
```

Db

RESULT 12

BU928625

```
LOCUS       BU928625                771 bp    mRNA    linear   EST 18-OCT-2002
DEFINITION  AGENCOURT_10421966 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:6654513
            5', mRNA sequence.
ACCESSION   BU928625
VERSION     BU928625.1  GI:24117355
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2906 row: e column: 09
            High quality sequence stop: 540.
FEATURES             Location/Qualifiers
     source          1..771
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6654513"
                     /tissue_type="glioblastoma"
                     /clone_lib="NIH_MGC_57"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTACGGCC-3', and 3' adaptor
                     sequence: 5'-ATTCTAGAGCGCGAGCGGCCGCGGGCGGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.55 Kb (range 0.9-4.0 kb). 12/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
```

US-09-701-618A-2 (1-141) x BU928625 (1-771)

```
Alignment Scores:
Pred. No.:
Score:                 715.00          Length:        771
Percent Similarity:    100.00%         Matches:       141
Best Local Similarity: 100.00%         Conservative:    0
Query Match:           5               Mismatches:      0
                                       Indels:          0
                                       Gaps:            0
```

ORIGIN

```
Qy    1 MetAlaGlyGluGluIleAsnGluAspProLeuGluGlnAlaLysValAspLeuLeuSerAla 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  ...

Qy  101 ThrAspLysGlyLysProLysLeuAspAlaArgGlyLysAlaSerArgPheValLysAsn 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  370 ACAGACAAGGAAAAGCCTAAGCTGGACGCACAGAGGAAGGCGAGCTTCAAGAGATTTGTAAAAAT 429

Qy  121 AlaLeuTrpGluProLysSerLysSerAsnAlaSerLysSerValAlaAsnLysGlyGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  430 GCCCTCTGGGAACCAAAATCGAAAAGTAACGCATCAAAAAGTGTCGCAATAAAAGGAGGAAAAAGTAAA 489

Qy  141 Ser 141
        |||
Db  490 AGT 492
```

Left column

```
Db   79  ATGGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAAATTCACGAGTATTGTCAGCG 138
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   21  PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
Db  139  TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGAGTGTCTGTTTCTAGA 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   41  AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60
Db  199  AATGAGTTGTTGCAGAAGTTGGATCCACTGAACAAGCAAAAGTGGATTGGTTCTGCA 258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61  TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
Db  259  TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81  HisProValLysGlnGluLeuGlnLysArgIleLeuArgValTyrMetAsnArgValLysGluIle 100
Db  319  CATCCAGTAAAACAGGAATTGGAAAAGAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  101  ThrAspLysLysLysAlaGlyAlaAlaSerArgPheValLysAsn 120
Db  379  ACAGACAAGAAAAAGGCTGGCAGAGCTGCACGCTTCAAGAATTTGTAAAAAAT 438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121  AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
Db  439  GCCCTCTGGGAACCAAAATCGAAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAAAGTAAA 498
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  141  Ser 141
         |||
Db  499  AGT 501
```

```
RESULT 13
CB998397
LOCUS        CB998397                789 bp    mRNA    linear   EST  01-MAY-2003
DEFINITION   AGENCOURT_13643706 NIH_MGC_187 Homo sapiens cDNA clone
             IMAGE:30320803 5', mRNA sequence.
ACCESSION    CB998397
VERSION      CB998397.1  GI:30292917
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 789)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Michael Bromstein and Dr. Miklos Palkovits
             cDNA Library Preparation: Dr. Michael Bromstein and Dr. Miklos Palkovits
             cDNA Library Arrayed by: CLONTECH Laboratories, Inc.
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: NDCM119 row: C column: 20
             High quality sequence stop: 545.
FEATURES             Location/Qualifiers
     source          1..789
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30320803"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_187"
                     /note="Organ: Blood vessels - aorta, basilar and artery;
                     Vector: pDNR-LIB; Site1: Sfi1 (ggccattatggcc); Site 2:
                     Sfi1 (ggccgcctcggcc); 5' and 3' adaptors were used in
                     cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3'; 3' adaptor sequence:
                     5'-ATTCTAGAGCGGCCGACGCCGGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
```

Right column

```
ORIGIN         (range 0.5-4.0 kb), 14/15 colonies contained inserts by
               PCR. This library was enriched for full-length clones and
               was constructed by Clontech Laboratories (Palo Alto, CA).
               Note: this is a NIH_MGC Library."

Alignment Scores:
  Pred. No.:             2.92e-80
  Score:                 715.00        Length:        789
  Percent Similarity:    100.00%       Matches:       141
  Best Local Similarity: 100.00%       Conservative:  0
  Query Match:           100.00%       Mismatches:    0
  DB:                    6             Indels:        0
                                       Gaps:          0

US-09-701-618A-2 (1-141) x CB998397 (1-789)
```

```
Qy    1  MetAlaGlyGluGluLeuAspLysLeuLysAspAlaGlyGlnAspTyrArgAlaGluMetTyrProValGlnIleHisGluTyrLeuSerAla 20
Db    1  ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAAATTCACGAGTATTGTCAGCG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   21  PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
Db   62  CATCCAGTAAAACAGGAATTGGAAAAGAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  101  ThrAspLysLysLysAlaGlyAlaAlaSerArgPheValLysAsn 120
Db  302  CATCCAGTAAAACAGGAATTGGAAAAGAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 361
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81  HisProValLysGlnGluLeuGlnLysArgIleLeuArgValTyrMetAsnArgValLysGluIle 100
Db  242  TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61  TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
Db  182  AATGAGTTGTTGCAGAAGTTGGATCCACTGAACAAGCAAAAGTGGATTGGTTCTGCA 241
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   41  AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla 60
Db  122  TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGAGTGTCTGTTTCTAGA 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   21  PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg 40
Db   62  ATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAAATTCACGAGTATTGTCAGCG 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121  AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
Db  422  GCCCTCTGGGAACCAAAATCGAAAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAAAGTAAA 481
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  141  Ser 141
         |||
Db  482  AGT 484
```

```
RESULT 14
BU568056
LOCUS        BU568056                813 bp    mRNA    linear   EST  16-SEP-2002
DEFINITION   AGENCOURT_10399045 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6614848
             5', mRNA sequence.
ACCESSION    BU568056
VERSION      BU568056.1  GI:22918356
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 813)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: CLONTECH
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
```

FEATURES
source
1. .813
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6614848"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_82"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3', and 3' adaptor sequence:
5'-ATTCTAGAGGGCCGCAGGCGGCCGACACGCGTCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

ORIGIN

US-09-701-618A-2 (1-141) x BU568056 (1-813)

Alignment Scores:
Pred. No.:
Score: 3.03e-80
Percent Similarity: 715.00   Length: 813
Best Local Similarity: 100.00%   Matches: 141
Query Match: 100.00%   Conservative: 0
DB: 5   Mismatches: 0
                                                Indels: 0
                                                Gaps: 0

Qy   1   MetAlaGlyGluGluGluIleAsnGluAspThrProValGluIleHisGluTyrLeuSerAla 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55   ATGGCAGGTGAAGAAGAAATTAATGAAGACATATCCAGTAGAAAATTCACGAGTATTTGTCAGCG 114

Qy   21   PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetSerValSerArg 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 115   TTTGAGAATTCAATTGGTGCTGTGGATGAGATGCTAAAGACCATGAGTGTCTGTTTCTAGA 174

Qy   41   AsnGluIleLeuGlnLysLeuArgIleArgValTyrMetAsnArgValLysGluIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 175   AATGAGTTGTTGCAGAAGCTGAGAATCCACTTGAACAGAGTGAAAAGTGGATTTGGTTTCGCA 234

Qy   61   TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 235   TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 294

Qy   81   HisProValLysGlnLeuGlnLeuArgIleArgValTyrMetAsnArgValLysGluIle 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 295   CATCCAGTAAAACAGGAATTCGAGAATCAGAGTAATATCAGAGAATTCACGAGTATTTGTCAGCG 354

Qy 101   ThrAspLysLysLysAlaGlyLysLysLysAspArgGlyAlaAlaSerArgPheValLysAsn 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 355   ACAGACAAGAAAAGGCTGCCAAGCTGCAGCTTCAAGATTTGTAAAAAAT 414

Qy 121   AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db 415   GCCCTCTCGGAACCAAAATCGATCAAAAAGTGCCAATAAAGGAAAAAGTAA 474

Qy 141   Ser 141
          |||
Db 475   AGT 477

FEATURES
source
1. .838
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6615564"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_82"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3', and 3' adaptor sequence:
5'-ATTCTAGAGGGCCGCAGGCGGCCGACACGCGTCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."
High quality sequence stop: 541.
Plate: LLCM2857 row: n column: 12
Location/Qualifiers

ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 838)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: CLONTECH
          cDNA Library Preparation: CLONTECH Laboratories, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          http://image.llnl.gov

FEATURES
source

US-09-701-618A-2 (1-141) x BU568457 (1-838)

Alignment Scores:
Pred. No.:
Score: 3.15e-80
Percent Similarity: 715.00   Length: 838
Best Local Similarity: 100.00%   Matches: 141
Query Match: 100.00%   Conservative: 0
DB: 5   Mismatches: 0
                                                Indels: 0
                                                Gaps: 0

Qy   1   MetAlaGlyGluGluGluIleAsnGluAspThrProValGluIleHisGluTyrLeuSerAla 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78   ATGGCAGGTGAAGAAGAAATTAATGAAGACTATCCAGTAGAAAATTCACGAGTATTTGTCAGCG 137

Qy   21   PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetSerValSerArg 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 138   TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTAAAGACAATGAGTGTCTGTTTCTAGA 197

Qy   41   AsnGluIleLeuGlnLysLeuArgIleArgValTyrMetAsnArgValLysGluIle 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 198   AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGAGCAAAAGTGGATTTGGTTTCGCG 257

Qy   61   TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 258   TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 317

Qy   81   HisProValLysGlnLeuGlnLeuArgIleArgValTyrMetAsnArgValLysGluIle 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 318   CATCCAGTAAAACAGGAATTCGAGAATCAGAGTAATATCAGAGAATTCACGAGTATTTGTCAGCG 377

Qy 101   ThrAspLysLysLysAlaGlyLysLysLysAspArgGlyAlaAlaSerArgPheValLysAsn 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 378   ACAGACAAGAAAAGGCTGCCAAGCTGCAGCAGGTCCAGCTTCAAGATTTGTAAAAAAAT 437

Qy 121   AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140

```
Db    438  GCCCTCTGGGACCAAAATCGAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAAAGTAAA  497
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    141  Ser                                                              141
               |||
Db    498  AGT                                                              500
```

This Page Blank (uspto)

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: January 5, 2005, 08:58:43 ; Search time 320.5 Seconds
(without alignments)
2309.418 Million cell updates/sec

Title: US-09-701-618A-2
Perfect Score: 715
Sequence: 1 MAGEEINEDYPVEIHEYLSA..........LWEPKSKNASKVANKGKSKS 141

Searched: 4134886 seqs, 2624710521 residues

Scoring table: BLOSUM62

		Xgapext	0.5
Xgapop	10.0 ;	Xgapext	0.5
Ygapop	10.0 ;	Ygapext	0.5
Fgapop	6.0 ;	Fgapext	7.0
Delop	6.0 ,	Delext	7.0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/USO970l618/runat_05012005_085838_24427/app_query.fasta_1.654
-DB=N_Geneseq_23Sep04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09701618 @CGN 1 1 886 @runat 05012005 085838 24427 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   N_Geneseq_23Sep04:*

Result No.	Score	% Query Match	Length DB	DB	ID	Description
1	715	100.0	1158	3	AAZ43927	Aaz43927 Human C1D
2	715	100.0	1172	6	ABL61981	Abl61981 Colon ade
3	715	100.0	1172	6	ABK84717	Abk84717 Human cDN
4	715	100.0	1172	10	ADF81400	Adf81400 Leukaemia
5	715	100.0	586	10	ADF81399	Adf81399 Leukaemia
6	699	97.8	586	12	ACH68585	Ach68585 Human gen

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

RESULT 1
AAZ43927

ID	AAZ43927	standard; cDNA; 1158 BP.
XX		
AC	AAZ43927;	
XX		
DT	17-MAR-2000 (first entry)	
XX		
DE	Human C1D cDNA.	
XX		
KW	C1D; human; apoptosis; tumour; gene therapy; treatment; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	118..543
FT		/*tag= a
FT		/product= "C1D"
XX		
PN	DE19824811-A1.	
XX		
PD	09-DEC-1999.	
XX		
PF	03-JUN-1998; 98DE-01024811.	
PR	03-JUN-1998; 98DE-01024811.	
XX		
PA	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.	
XX		
PI	Rothbarth K, Stammer H, Werner D;	

ALIGNMENTS

c	7	671	93.8	426	6	AAL43974	Aal43974 Mammalian
c	8	671	93.8	426	6	AAL43973	Aal43973 Mammalian
c	9	667	93.3	420	6	ADI28090	Adi28090 Human nuc
c	10	667	93.3	420	6	ADI28091	Adi28091 Human nuc
c	11	654	91.5	426	12	ADJ92841	Adj92841 Human co-
c	12	646	90.3	1038	3	AAZ43928	Aaz43928 Murine C1
c	13	644	90.1	420	6	AAL43975	Aal43975 Mammalian
c	14	644	90.1	420	6	AAL43976	Aal43976 Mammalian
c	15	519	72.6	380	3	AAC01237	Aac01237 Human sec
c	16	266	37.2	204	12	ACH82285	Ach82285 Human gen
c	17	150.5	21.0	1000	3	AAC49672	Aac49672 Arabidops
c	18	150.5	21.0	637	4	AAC40266	Aac40266 Arabidops
19	135	18.9	699	12	ADP98823	Adp98823 C. albica	
20	115	16.1	555	10	ACC61644	Acc61644 Gene sequ	
21	115	16.1	555	10	ADK64283	Adk64283 Disease t	
22	110	15.4	65	6	ABN53492	Abn53492 Mouse spl	
23	106	14.8	65	6	ABL14530	Abl14530 Drosophil	
24	103	14.4	2705	4	ABL14531	Abl14531 Drosophil	
25	96	13.4	110000	10	ACF67367_12	Continuation (13 o	
c	26	96	13.4	182624	6	ACF65379	Acf65379 Photorhab
c	27	95	13.3	60	6	ABN34297	Abn34297 Human spl
c	28	89	12.4	110000	10	ACF67367_43	Continuation (44 o
c	29	89	12.4	110000	10	ACF65388_04	Continuation (5 of
c	30	88.5	12.4	770	10	ACF66303	Acf66303 Photorhab
c	31	88.5	12.4	110000	10	ACF67367_13	Continuation (14 o
c	32	88.5	12.4	110000	10	ACF67367_12	Continuation (13 o
c	33	88.5	12.4	182624	6	ACF65379	Acf65379 Photorhab
c	34	88	12.3	110000	10	ACF65379_40	Continuation (41 o
c	35	88	12.3	110000	10	ACF67367_07	Continuation (8 of
c	36	86.5	12.1	110000	10	ACF65388_07	Continuation (8 of
c	37	86.5	12.1	110000	10	ACF67367_25	Continuation (26 o
c	38	86.5	12.1	110000	10	ACF65386_3	Continuation (4 of
c	39	85	11.9	456	10	ADH85189	Adh85189 Enterococ
40	85	11.9	3885	4	ABL13371	Abl13371 Drosophil	
41	85	11.9	8751	4	ABL13370	Abl13370 Drosophil	
42	83.5	11.7	1767	8	ACA33095	Aca33095 Prokaryot	
43	83.5	11.7	1785	10	ADH83717	Adh83717 Enterococ	
44	83.5	11.7	24601	2	AAX11160	Aax11160 Enterococ	
45	83.5	11.7	24601	2	ABS98955	Abs98955 Enterococ	
45	83	11.6	3747	8	ACA48336	Aca48336 Prokaryot	

```
XX
DR   WPI; 2000-063506/06.
DR   P-PSDB; AAY51024.
XX
PT   Inducing apoptosis by overexpressing the CID gene, particularly for
PT   treating tumors.
XX
PS   Claim 3; Fig 1; 10pp; German.
XX
CC   This invention describes a novel method for inducing apoptosis which
CC   comprises overexpressing the CID gene (I). The method is particularly
CC   used to treat tumors and can also be used in gene therapy. The method has
CC   no side effects on normal cells (contrast known methods of inducing
CC   apoptosis such as cytotoxins and radiation)', and may be effective on
CC   cells resistant to conventional treatments. Overexpression of (I) is
CC   sufficient to induce apoptosis but the effect may be increased
CC   when used in combination with other anti-tumor methods. When cells
CC   transfected with (I) undergo apoptosis, they release factors that kill
CC   neighboring, non-transfected cells (bystander effect). This sequence
CC   encodes the human CID protein described in the method of the invention
XX
SQ   Sequence 1158 BP; 388 A; 155 C; 237 G; 378 T; 0 U; 0 Other;

US-09-701-618A-2 (1-141) x AAZ43927 (1-1158)

Alignment Scores:
Pred. No.:
Score:                  8.98e-79
Percent Similarity:     100.00%      Length:           1158
Best Local Similarity:  100.00%      Matches:          141
Query Match:            100.00%      Conservative:     0
DB:                     3            Mismatches:       0
                                     Indels:           0
                                     Gaps:             0

Qy    1 MetAlaGlyGluGluGlnIleAsnGluAspTyrProValGlyIleHisGluTyrLeuSerAla 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  118 ATGGCAGGTGAAGAAAATTAATGAAGACTATCCAGTAGGAAATTCACCGAGTATTTGTCAGCG 177

Qy   21 PheGluAsnSerIleGlyAlaValAlaSpGluMetLeuLysThrMetSerValSerArg 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  178 TTTGAGAATTCCATTGGCGTCGTGGATGAGATGCTGAAGACCATGAGTGTCGTTTCTAGA 237

Qy   41 AsnGluLeuLeuGlnIleLysGlyLeuAspProLeuGluGlnIleAlaLysValAlaSerSerAla 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  238 AATGAGTTGTTGCAGAGTTGGATCCCACTTGAACAAGCAAAAGTGGAATTGGGTTTCTGCA 297

Qy   61 TyrThrIleAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  298 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357

Qy   81 HisProValLysGlnGluLeuGlnLeuArgIleArgValTyrMetAsnArgValLysGluIle 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  358 CATCCAGTAAAAACAGGAATTCGAAAGATCGAGTATATGAACAGAGCGTCAAGGAAATA 417

Qy  101 ThrAspLysLysAlaGlyLysAlaLysLysAlaSerArgPheValLysAsn 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  418 ACAGACAAGAAAAGGCTCGGCAACGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAT 477

Qy  121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  478 GCCCTCTGGGAACCAAAAATCGAAAAATGCCAATAAAGGAAAAAAGTAA 537

Qy  141 Ser 141
          |||
Db  538 AGT 540
```

DE   Colon adenocarcinoma related gene sequence SEQ ID NO:318.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; dž.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0205531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0251334P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0345509P.
PR   22-SEP-2000; 2000US-0256388P.
PR   26-SEP-2000; 2000US-0256388P.
PR   26-SEP-2000; 2000US-0256377P.
PR   27-SEP-2000; 2000US-0257811P.
PR   27-SEP-2000; 2000US-0345667P.
PR   27-SEP-2000; 2000US-0357720P.
PR   25-SEP-2000; 2000US-0344923P.
PR   25-SEP-2000; 2000US-0249024P.
PR   28-SEP-2000; 2000US-0235077P.
PR   28-SEP-2000; 2000US-0235082P.
PR   28-SEP-2000; 2000US-0235840P.
PR   28-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0260033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   29-SEP-2000; 2000US-0261109P.
PR   29-SEP-2000; 2000US-0261111P.
PR   29-SEP-2000; 2000US-0236891P.
PR   29-SEP-2000; 2000US-0236842P.
PR   02-OCT-2000; 2000US-0271172P.
PR   02-OCT-2000; 2000US-0271173P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0372795P.
PR   03-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237599P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene or a signature gene set.
XX
PS   Claim 1; SEQ ID NO 318; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
```

```
RESULT 2
ID   ABL61981 standard; DNA; 1172 BP.
XX
AC   ABL61981;
XX
DT   15-MAY-2002 (first entry)
XX
```

```
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (1) of a signature gene set, where (1)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (1) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;

US-09-701-618A-2 (1-141) x ABL61981 (1-1172)

Alignment Scores:
Pred. No.:                           Length:         1172
Score:              9.13e-79         Matches:        141
Percent Similarity: 715.00           Conservative:   0
Best Local Similarity: 100.00%       Mismatches:     0
Query Match:        100.00%          Indels:         0
DB:                 6                Gaps:           0


Qy     1  MetAlaGlyGluGluIleAsnGluAspThrProValGluIleHisGluThrTyrLeuSerAla  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   118  ATGGCAGGTGAAGAAATTAATGAAGACAGTACTATCCAGTGAGAAATTCACGAGTATTGTCAGCG 177

Qy    21  PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178  TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 237

Qy    41  AsnGluLeuLeuGlnLysGlyLeuAspProLeuGlnGlnAlaLysValLysValSerArgAla  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   238  AATGAGTTGTTGCAGAAGTTGGATCCACTTCAACAAGCAAAAGTGAAAGTGAGTCGTGCA 297

Qy    61  TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   298  TACACATTAAATTCAAATGTTTTGGGTTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357

Qy    81  HisProValLysGlnGlnLeuGluGluValThrTyrMetAsnArgValLysGluIle  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   358  CATCCAGTAAAACAGGAATTGGAAGAAGTCACAGTATATATGAACAGAGTCAAGGAGAAATA 417

Qy   101  ThrAspLysLysLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   418  ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCAGCTTCAAGATTTGTAAAAAAT 477

Qy   121  AlaLeuTrpGluProLysSerLysAsnAlaSerArgValAlaAsnLysGlyLysSerLys  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   478  GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAGTAAA 537

Qy   141  Ser  141
          |||
Db   538  AGT  540


RESULT 3
ABK84717
ID     ABK84717 standard; cDNA; 1172 BP.
XX
AC     ABK84717;
XX
DT     14-AUG-2002  (first entry)
XX
DE     Human cDNA differentially expressed in granulocytic cells #1288.
XX
KW     Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW     viral infection; parasitic infection; protozoal infection;
KW     fungal infection; sterile inflammatory disease; psoriasis;
KW     rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW     cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW     adult respiratory distress syndrome; inflammatory bowel disease;
KW     Crohn's disease; ulcerative colitis; periodontal disease;
KW     granulocyte activation; chronic inflammation; allergy.
XX
OS     Homo sapiens.
XX
PN     WO20228999-A2.
XX
PI     11-APR-2002.
XX
DR     WPI; 2002-435328/46.
XX
PT     03-OCT-2001; 2001WO-US030821.
XX
PT     03-OCT-2000; 2000US-0237189P.
XX
PA     (GENE-) GENE LOGIC INC.
XX
PI     Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
PS     Claim 1; SEQ ID NO 1288; 114pp; English.
XX
CC     The invention relates to detecting (M1) granulocyte (GC) activation
CC     (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC     DNA chip analysis as given in the specification, and comparing the
CC     expression level to an expression level in an unactivated GC, where
CC     differential expression of Gs is indicative of GCA. Also included are
CC     modulating (M2) GA by contacting GC with an agent that alters the
CC     expression of at least one gene in Gs; (2) screening (M3) for an agent
CC     capable of modulating GCA or an inflammation (especially chronic) in a
CC     tissue, an allergic response in a subject, exposure of a subject to a
CC     pathogen or sterile inflammatory disease using the gene expression
CC     profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC     tissue, an allergic response in a subject, exposure of a subject to a
CC     pathogen or sterile inflammatory disease, by detecting the level of
CC     expression in a sample of the gene (s) from Gs, where the level of
CC     of expression of the gene is indicative of inflammation; (4) treating
CC     (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC     response in a subject, exposure of a subject to a pathogen or sterile
CC     inflammatory disease, by contacting a tissue having inflammation with an
CC     agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC     is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC     for screening an agent capable of modulating GCA preferably in an
CC     inflammation in a tissue; M4 is useful for detecting an inflammation
CC     (especially chronic) in a tissue, an allergic response in a subject,
CC     exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC     psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC     cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC     respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC     disease, ulcerative colitis, periodontal disease; also bacterial
CC     infection, viral infection, parasitic infection, protozoal infection,
CC     fungal infection and M5 is useful for treating one of the above
CC     conditions. The present sequence represents a gene differentially
CC     expressed in granulocytes. Note: The sequence data for this patent did
CC     not form part of the printed specification, but was obtained in
CC     electronic format directly from WIPO at
CC     ftp.wipo.int/pub/published_pct_sequences
XX
SQ     Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                           Length:         1172
Score:              9.13e-79         Matches:        141
Percent Similarity: 715.00           Conservative:   0
Best Local Similarity: 100.00%       Mismatches:     0
Query Match:        100.00%          Indels:         0
DB:                 6                Gaps:           0
```

Detecting granulocyte activation by detecting differential expression of
genes associated with granulocyte activation, which serves as diagnostic
markers that is useful for monitoring disease states and drug toxicity.

US-09-701-618A-2 (1-141) x ABK84717 (1-1172)

```
Qy    1 MetAlaGlyGlyGlyGlnIleAsnGluAspTyrProValGlnIleHisGluTyrLeuSerAla 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  118 ATGGCAGGTGGAGGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 177

Qy   21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  178 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 237

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnAlaLysValAspLeuValSerAla 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  238 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 297

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  298 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357

Qy   81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  358 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 417

Qy  101 ThrAspLysLysLysAlaGlyLysAlaLeuAspSerArgPheValLysAsn 120
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  418 ACAGACAAGAAAAAGGCTGGCAAAGCTTCAGATTTCGTAAAGAAT 477

Qy  121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  478 GCCCTCTGGGAACCCAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAA 537

Qy  141 Ser 141
        |||
Db  538 AGT 540
```

```
XX
PS    Disclosure; SEQ ID NO 1956; 2938pp; English.
XX
CC    The present invention relates to a method (M1) for determining the
CC    subtype of leukaemia cells and whether a patient sample contains
CC    leukaemia cells. The method comprises determining the expression profile
CC    of a group of markers in a patient sample. The method is useful for
CC    determining the presence of leukaemia cells, its types or subtypes, and
CC    for the preparation of a medicament for treating leukaemia.
XX
SQ    Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:
Score:                  9.13e-79
Percent Similarity:     715.00
Best Local Similarity:  100.00%
Query Match:            100.00%
DB:                     10

Length:        1172
Matches:       1172
Conservative:  141
Mismatches:    0
Indels:        0
Gaps:          0
```

RESULT 4
ID   ADF81400
AC   ADF81400;
XX
DT   26-FEB-2004 (first entry)
XX
DE   Leukaemia-related DNA sequence #1956.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
OS   Unidentified.
PN   WO200303943-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.
PR   30-APR-2002; 2002EP-0009758.
PR   05-NOV-2001; 2001EP-0012644.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
PA   (SCHO/) SCHOCH C.
PA   (KERN/) KERN W.
XX
PI   Haferlach T, Schoch C, Kern W, Schnittger S, Dugas M;
PI   Eils R, Brors B, Mergenthaler S;
XX
DR   WPI; 2003-505037/47.
XX
PT   Determining the subtype of leukemia cells and whether a patient sample
PT   contains leukemia cells and whether a patient sample
PT   comprises determining the expression profile of a group of markers in a
PT   patient sample.

US-09-701-618A-2 (1-141) x ADF81400 (1-1172)

```
Qy    1 MetAlaGlyGlyGlyGlnIleAsnGluAspTyrProValGlnIleHisGluTyrLeuSerAla 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  118 ATGGCAGGTGGAGGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG 177

Qy   21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  178 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 237

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnAlaLysValAspLeuValSerAla 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  238 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 297

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  298 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357

Qy   81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  358 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 417

Qy  101 ThrAspLysLysLysAlaGlyLysAlaLeuAspSerArgPheValLysAsn 120
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  418 ACAGACAAGAAAAAGGCTGGCAAAGCTTCAGATTTCGTAAAGAAT 477

Qy  121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  478 GCCCTCTGGGAACCCAAAAATGCATCAAAAGTTGCCAATAAAGGAAAAAGTAAA 537

Qy  141 Ser 141
        |||
Db  538 AGT 540
```

RESULT 5
ID   ADF81399
AC   ADF81399;
XX
DT   26-FEB-2004 (first entry)
XX
DE   Leukaemia-related DNA sequence #1955.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
OS   Unidentified.
XX
PN   WO200303943-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.

ADF81399 standard; DNA; 1172 BP.

```
XX
PR   05-NOV-2001; 2001EP-00126244.
PR   30-APR-2002; 2002EP-00009758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
PA   (SCHO/) SCHOCH C.
PA   (KERN/) KERN W.
XX
PI   Haferlach T, Schoch C, Kern W, Schnittger S, Dugas M;
PI   Ells R, Brors B, Mergenthaler S;
XX
DR   WPI; 2003-505037/47.
XX
PS   Disclosure; SEQ ID NO 1955; 2938pp; English.
XX
PT   Determining the subtype of leukemia cells and whether a patient sample
PT   contains leukemia cells or other cells, useful for treating leukemia,
PT   comprises determining the expression profile of a group of markers in a
PT   patient sample.
XX
CC   The present invention relates to a method (M1) for determining the
CC   subtype of leukaemia cells and whether a patient sample contains
CC   leukaemia cells. The method comprises determining the expression profile
CC   of a group of markers in a patient sample. The method is useful for
CC   determining the presence of leukaemia cells, its types or subtypes, and
CC   for the preparation of a medicament for treating leukaemia.
XX
SQ   Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;

Alignment Scores:
Pred. No. :
Score:                  715.00          Length:        1172
Percent Similarity:     100.00%         Matches:       141
Best Local Similarity:  100.00%         Conservative:  0
Query Match:            100.00%         Mismatches:    0
DB:                     10              Indels:        0
                                        Gaps:          0

US-09-701-618A-2 (1-141) x ADF81399 (1-1172)

Qy   1   MetAlaGlyGluGlyIleAsnGluAsnTyrProValGluIleHisGluTyrLeuSerAla 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   118 ATGGCAGGTGAAGGAAATTAATGAAGAACTACCCAGTAGAAATTCACGAGTATTTGTCAGCG 177

Qy   21  PheGluAsnSerIleCysIleAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178 TTTGAGAATTCCATTGGTGTCGTGGATGAGATGCTGAAGACCATGATGTCTGTTTCTAGA 237

Qy   41  AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAlaSerLeuValSerArgAla 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   238 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTCGCA 297

Qy   61  TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnValTyrAsnProLysGlu 80
         ||||||||||||||||||||||||||||||||||||||||||||
Db   298 TACACATTAAATTCAAATGTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA 357

Qy   81  HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGlyIle 100
         |||||||||||||||||||||||||||||||||||||||||||||
Db   358 CATCCAGTAAAACAGGAATTGGAAAGAATCAGAGTTATATGAACAGAGTCAAGGAAATA 417

Qy   101 ThrAspLysGlyLysAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
         |||||||||||||||||||||||||||||||||||||||||
Db   418 ACAGACAAGGGAAAAGGCTGGCAAGCTGGACAGAGGTGGCAGCTTCAAGATTTGTAAAAAAT 477

Qy   121 AlaLeuTrpGluProLysGlnLysSerLysAsnAlaAsnLysSerLysTyrLysSerLys 140
         ||||||||||||||||||||||||||||||||||||||||
Db   478 GCCCTCTGGGAACCAAAATCGAAAAATGCATCAAAAAAGTTGCCAATAAAGGAAAAAGTAAA 537

Qy   141 Ser 141
         |||
Db   538 AGT 540
```

```
RESULT 6
ACH68585/c
ID   ACH68585 standard; DNA; 586 BP.
XX
AC   ACH68585;
XX
DT   29-JUL-2004 (first entry)
XX
PN   US2003194704-A1.
XX
OS   Homo sapiens.
XX
DE   Human genome derived single exon probe #1780.
XX
KW   Human; probe; ss; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG, Rank DR, Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
XX
PS   Claim 15; SEQ ID NO 1780; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately
CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, a vector comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   tissues to detect the presence of expressed messages that contain their
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in identifying and characterising
CC   alternative splicing events, in detecting and characterising gross
CC   alterations in the genomic locus that includes their exon, in assessing
CC   smaller genomic alterations or in priming the synthesis of nucleic acids,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=2003194704
```

XX
SQ   Sequence 586 BP; 145 A; 138 C; 109 G; 194 T; 0 U; 0 Other;
XX

Alignment Scores:
Pred. No.:
Score: 3.46e-77
Percent Similarity: 69.00
Best Local Similarity: 98.57%
Query Match: 98.57%
                                97.76%

                     Length: 586
                     Matches: 138
                     Conservative: 0
                     Mismatches: 2
                     Indels: 0
                     Gaps: 0
DB:

US-09-701-618A-2 (1-141) x ACH68585 (1-586)

```
Qy    2 AlaGlyGluGluLeuAsnGluAspTyrProValGlyIleLeuHisGluTyrLeuSerAlaPhe 21
Db  515 GCAGGTGAAGAAATTAATGAAGAACATATCCAGTAGAAAATTCACGAGTAATTGTCAGCGTTT 456

Qy   22 GluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetMetSerValSerArgAsn 41
Db  455 GAGAATTCCAATTGGTCTGCTGTGGATCGAGATGCTGAGAACGTCTGTTTCTGTTCTTAGAAAT 396

Qy   42 GluLeuLeuGlnLysLeuAspProLeuGlnAlaLysValSerAlaThrTyr 61
Db  395 GAGTTGTTGCAGAAGTTGGATCCACTTGAACAGGCAAAAGTGGAATTTGGTTCGTTCTCATAC 336

Qy   62 ThrLeuAsnSerMetPheThrProValTyrLeuAlaThrArgGlnGlyIleValAsnProLysGluHis 81
Db  335 ACATTAAATTCAATGTTTGGGTTTATTGGCCAACCACTTGAACAGGAGTTAATCCTAAGGAACAT 276

Qy   82 ProValLysGlyIleGluLeuArgIleArgValTyrMetAsnArgValLysGlyIleThr 101
Db  275 CCAGTAAAACAGGAATTCAAATGTTTGGGTTTATTGGCCAGAGTATATGAATAGAGTCAAAGGAAATAACA 216

Qy  102 AspLysLysGlyLysAlaGlyLysLeuAspArgGlyValAlaAlaSerArgPheValLysAsnAla 121
Db  215 GACAAGAAAAAGGCTGGCAAGCTCGACAGAGGGTGCAGCTTCAAGATTTGTGTTAAAAATGCC 156

Qy  122 LeuTrpGluProLysSerLysSerValAlaAsnLysGlyLysSerLysSer 141
Db  155 CTCTGGGAACCAAAAACCAAAAATGCATCCAAAAGTTGCCAATAAAGGAAAAAGTAAAAGT 96
```

RESULT 7
AAL43974/c
ID   AAL43974 standard; cDNA; 426 BP.
AC   AAL43974;
XX
DT   27-SEP-2002 (first entry)
XX
DE   Mammalian nuclear receptor cofactor CF7 reverse complement sequence.
XX
KW   Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;
KW   metabolism regulation; cell homeostasis; cell proliferation;
KW   differentiation; pathological cellular aberration;
KW   cellular defence mechanism.
XX
OS   Mammalia.
XX
PN   WO200242322-A2.
PD   30-MAY-2002.
PF   21-NOV-2001; 2001WO-EP013548.
PR   21-NOV-2000; 2000EP-00125524.
PA   (LION-) LION BIOSCIENCE AG.
PI   Jackson D, Casari G, Suckow J;
DR   WPI; 2002-565559/60.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.
XX
PS   Claim 1; Fig 2; 68pp; English.
XX

SQ   Sequence 426 BP; 105 A; 99 C; 64 G; 158 T; 0 U; 0 Other;
XX

CC   The invention comprises the amino acid and coding sequences of two
CC   mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC   CF8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CF7
CC   and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC   involved in cellular functions, such as: regulation of metabolism and
CC   cell homeostasis; cell proliferation and differentiation, pathological
CC   cellular aberrations, or cellular defence mechanisms. The present cDNA
CC   sequence represents the reverse complement of the mammalian nuclear
CC   receptor cofactor CF7 gene sequence

US-09-701-618A-2 (1-141) x AAL43974 (1-426)

Alignment Scores:
Pred. No.:
Score: 6.76e-74
Percent Similarity: 671.00
Best Local Similarity: 95.74%
Query Match: 93.62%
                                93.85%

                     Length: 426
                     Matches: 132
                     Conservative: 3
                     Mismatches: 6
                     Indels: 0
                     Gaps: 0
DB:

```
Qy    1 MetAlaGlyGluGluLeuAsnGluAspTyrProValGlyIleLeuHisGluTyrLeuSerAla 20
Db  426 ATGGCAGGTGAAGAAATTAATGAAGAACATATCCAGTAGAAAATTCACGAGTAATTGTCAGCA 367

Qy   21 PheGluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetMetSerValSerArg 40
Db  366 TTTGGAGAATTCCAATTGATCGTGTGGATCGAGATGCTGAGAACGTCTGTTTCTCGTAGA 307

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnAlaLysValSerAlaLeuAlaSerAla 60
Db  306 AATGAGTTGTTGCAGAAGTTGGACCCACTTGAACAGGCAAAAGTGGAATCCTAAGGCA 247

Qy   61 TyrThrThrLeuAsnSerMetPheThrProValTyrLeuAlaThrArgGlnGlyIleValAspAsnProLysGlu 80
Db  246 TACACATTAAATTCAATGTTTGGGTTTATTGGCCAACCACTTGAACAGGAGTTAATCCTAAGGAA 187

Qy   81 HisProValLysGlyIleGluLeuArgIleArgValLysGlyIle 100
Db  186 CATCCAGTAAAGCAGGAATTCGAAAGAATTCAGAGTATATGAATA 127

Qy  101 ThrAspLysLysGlyLysAlaGlyLysLeuAspArgGlyValAlaAlaSerArgPheValLysAsn 120
Db  126 ACAGACAAGAAAAAGGCTGGCAAGCTCGACAGAGGGTGCAGCTTCAAGATTTGTAAGAAAT 67

Qy  121 AlaLeuTrpGluProLysSerLysSerValAlaAsnLysGlyLysSerLys 140
Db   66 GCCCTCTGGGAACCAAAAACCAAAAATACATCCAAAAGTTGCCCAATAAAGGAAAAGTAAA 7

Qy  141 Ser 141
        |||
Db    6 AGT 4
```

RESULT 8
AAL43973
ID   AAL43973 standard; cDNA; 426 BP.
AC   AAL43973;
XX
DT   27-SEP-2002 (first entry)
XX
DE   Mammalian nuclear receptor cofactor CF7 coding sequence.
XX
KW   Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;
KW   metabolism regulation; cell homeostasis; cell proliferation;
KW   differentiation; pathological cellular aberration;

```
KW   cellular defence mechanism.
XX
OS   Mammalia.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..426
FT                   /*tag= a
FT                   /product= "Mammalian CP7 protein"
XX
PN   WO20024232322-A2.
XX
PD   30-MAY-2002.
XX
PF   21-NOV-2000; 2000EP-00125524.
PR   21-NOV-2001; 2001WO-EP013548.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Jackson D, Casari G, Suckow J;
XX
DR   WPI; 2002-565559/60.
DR   P-PSDB; AA015405.
XX
PT   Novel nuclear receptor cofactors, CP7 and CP8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.
XX
PS   Claim 1; Pig 1; 68pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of two
CC   mammalian nuclear receptor cofactor proteins (CP7 and CP8). The CP7 and
CC   CP8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CP7
CC   and/or CP8. The CP7 and CP8 DNA sequences are useful for modulating genes
CC   involved in cellular functions, such as: regulation of metabolism and
CC   cell homeostasis, cell proliferation and differentiation, pathological
CC   cellular aberrations, or cellular defence mechanisms. The present cDNA
CC   sequence encodes the mammalian nuclear receptor cofactor CP7 protein
XX
SQ   Sequence 426 BP; 158 A; 64 C; 99 G; 105 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                  6.766-74
Score:                  671.00          Length:           426
Percent Similarity:     95.74%          Matches:          132
Conservative:           3               Conservative:     3
Best Local Similarity:  93.62%          Mismatches:       6
Query Match:            93.85%          Indels:           0
DB:                     6               Gaps:             0

US-09-701-618A-2 (1-141) x AAL43973 (1-426)

Qy    1 MetAlaGlyGluGluIleAsnGluAspThrProValTyrLeuSerAla 20
          ||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGCAGGTGAAGAAATTAATGAGGACTATCCAGTAGAAATTCACGATTATTTGTCAGCA 60

Qy   21 PheGluAsnSerIleGlyAlaValAlaValAspGluMetLeuLysThrMetSerValSerArg 40
          |||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTTGCGAATTCCATTGATGCTGTGGCGATGAGAATGCTGAAGAACACATGATGTCTGTTTCTAGA 120

Qy   41 AsnGluLeuLeuGlnLysValGluAspProLeuGluGlnGlnAlaLysValAspLeuValSerAla 60
          |||||||||||||||||||||||||||||||||||||||||||
Db  121 AATGAAGTTGTTGCAGAAGTTGGACCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 180

Qy   61 TyrThrLeuAsnSerMetPheThrProValTyrLeuAlaThrGlnProLysGlu 80
          ||||||||||||||||||||||||||||||||||||||||
Db  181 TACACATTAAATTCAATGTTTTGGCAACTCAAGGAGTGTATCCTAAGGAA 240

Qy   81 HisProValLysGlnLeuGluArgValTyrMetAsnArgValLysGluIle 100
          |||||||||||||||||||||||||||||||||||||||
Db  241 CATCCAGTAAAGCAAGGAGAATTGGAAAGAGTATATATGAACAGAGTCAAGGAAATA 300

Qy  101 ThrAspLysLysValAlaGlyLysLeuAspArgGlyAlaAlaSerArgPheValLysAsn 120
          ||||||||||||||||||||||||||||||||||||||||
Db  301 ACAGACAAGAAAGTGGCTGGGAAGCTGGACAGAGGTGCAGCTTCAAGATTGTAAGAAAT 360

Qy  121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
          ::|||||||||||||||||| |||| ||||||||||||||||||::|||
Db  361 GCCCTCTGGGAACCAAAAAGCAAAATACAATCCAAAAGTTGCCATAAAAGGAAAAAGTAAA 420

Qy  141 Ser 141
          |||
Db  421 AGT 423

RESULT 9
ADI28090
ID   ADI28090 standard; DNA; 420 BP.
XX
AC   ADI28090;
XX
DT   22-APR-2004   (first entry)
XX
DE   Human nuclear receptor cofactor CP6 sense coding strand.
XX
KW   gene; nuclear receptor cofactor; CP6; cellular function inhibition;
KW   metabolic disorder; immunological indication; hormonal dysfunction;
KW   neurosystemic disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..420
FT                   /*tag= a
XX
PN   WO200224728-A2.
XX
PD   28-MAR-2002.
XX
PF   17-SEP-2001; 2001WO-EP010744.
PR   22-SEP-2000; 2000EP-00120722.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Casari G, Jackson D;
XX
DR   WPI; 2002-383179/41.
DR   P-PSDB; ADI28083.
XX
PT   Polynucleotide and polypeptide of novel nuclear receptor cofactor useful
PT   for screening drugs regulating cofactor-associated physiological
PT   responses e.g. hormonal dysfunctions.
XX
PS   Claim 1; SEQ ID NO 1; 97pp; English.
XX
CC   The invention relates to an isolated polynucleotide encoding a nuclear
CC   receptor cofactor, also known as CP6. The polynucleotide or encoded
CC   protein is useful for construction of multiple nuclear receptor cofactor
CC   specific sequence alignments. The protein is useful for the construction of protein
CC   sequence alignments. The protein is useful for the construction of protein
CC   of inhibiting the cellular function of the cofactor CP6. The
CC   polynucleotide is useful for making vectors and for transforming cells,
CC   both of which are ultimately useful for production of the CP6 protein.
CC   They are also useful as scientific research tools for developing nucleic
CC   acid probes for determining expression levels of the cofactor gene, e.g.
CC   to identify diseased or otherwise abnormal states. They are particularly
CC   useful for diagnostic purposes to e.g. identify deleted or mutant CP6
CC   genes; or their measure expression. They are useful for developing
CC   analytical tools such as antisense oligonucleotide for selectively
CC   inhibiting expression of the cofactor gene to determine physiological
CC   responses. The protein is useful for screening drugs for agonist and
CC   antagonist activity, and therefore, for screening drugs useful in
CC   regulating physiological responses associated with the cofactors such as
CC   metabolic disorders, immunological indications, hormonal dysfunction,
CC   neurosystemic diseases. The proteins are also useful for developing
CC   antibodies for detection of the proteins. The polynucleotide can be used
```

Left column

```
CC    to design primers for a polymerase chain reaction and are also used to
CC    model the three-dimensional structure of the protein. This sequence
CC    represents the sense strand of the CF6 gene coding region. (Note: this
CC    sequence appears to be a fragment of the CF6 gene
CC    from the CF6 gene sequence given in Fig 1 (also designated SEQ ID NO: 1)
CC    of the specification).
XX
SQ    Sequence 420 BP; 155 A; 64 C; 98 G; 103 T; 0 U; 0 Other;
```

US-09-701-618A-2 (1-141) x ADI28090 (1-420)

Alignment Scores:

Pred. No.:			
Score:	2.08e-73		
Percent Similarity:	95.71%	Length:	420
Percent Local Similarity:	93.57%	Matches:	131
Best Local Similarity:		Conservative:	3
Query Match:	93.29%	Mismatches:	6
DB:	6	Indels:	0
		Gaps:	0

```
  1 MetAlaGlyGluGluIleLeuAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
    ||||||||||||||||||||||||||||||::|||||||||||||||||||||||
  1 ATGGCAGGTGAAGAAATTAATGAGGACTATCCAGTAGAAATTCACGAATATTTGTCAGCA 60

 21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
    |||
 61 TTTGCGAATTCCATTGATGGTGTGGACATGCTGAAGACAATGATGTCTGTTTCGTAGA 120

 41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAlaSerValSerAla 60
    ||||||||||||||||||||||||||||||||||||||
121 AATGAGTTGTTGCAGAAGTTGGACCCACTGGAACAAGCAAAAGTGGCATCCGTTTCTGCA 180

 61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
    |||||||||||||||||||||||||||||||||||||
181 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAGGAGTGAATCCTAAGGA 240

 81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
    ||||||||||||||||||||||||||||||||||||
241 CATCCAGTAAAGCAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 300

101 ThrAspLysLysLysAlaLeuGlyLysLysLeuAspAspAlaSerArgPheValLysAsn 120
    ||||||||||||||||||||||||::|||||||::|||
301 ACAGACAAGAAAAAAGGCTGGCAAGCTCGGACAGAGGTGCAGCTTCAAGATTTGTAAGAAAT 360

121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
    |||||||||||||||||::||||::|||
361 GCCCTCTGGGAACCAAAATCCAAAGTTGCCCATAAAGGAAAAAGTAAA 420
```

RESULT 10
```
ID   ADI28091/c
XX
AC   ADI28091;
XX
DT   22-APR-2004 (first entry)
XX
DE   Human nuclear receptor cofactor CF6 coding sequence complement strand.
XX
KW   neurosystemic disease; ss.
KW   metabolic disorder; immunological disease; hormonal dysfunction;
KW   gene; nuclear receptor cofactor CF6; cellular function inhibition;
XX
OS   Homo sapiens.
XX
PN   WO20224728-A2.
XX
PD   28-MAR-2002.
XX
PF   17-SEP-2001; 2001WO-EP010744.
PR   22-SEP-2000; 2000EP-00120722.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
```

Right column

```
PI   Casari G, Jackson D;
XX
DR   WPI; 2002-383179/41.
XX
PT   Polynucleotide and polypeptide of novel nuclear receptor cofactor useful
PT   for screening drugs regulating cofactor-associated physiological
PT   responses e.g. hormonal dysfunctions.
XX
PS   Claim 1; SEQ ID NO 2; 97pp; English.
XX
CC   The invention relates to an isolated polynucleotide encoding a nuclear
CC   receptor cofactor, also known as CF6. The polynucleotide or encoded
CC   protein is useful for construction of multiple nuclear receptor cofactor
CC   specific sequence alignments, preferably for the construction of protein
CC   sequence alignments. The protein is useful for screening agents capable
CC   of inhibiting the cellular function of the cofactor CF6. The
CC   polynucleotide is useful for making vectors and for transforming cells,
CC   both of which are ultimately useful for production of the CF6 protein.
CC   They are also useful as scientific research tools for developing nucleic
CC   acid probes for determining expression levels of the cofactor gene, e.g.
CC   to identify diseased or otherwise abnormal states. They are particularly
CC   useful for diagnostic purposes to e.g. identify deleted or mutant CF6
CC   genes; or their measure expression. They are useful for developing
CC   analytical tools such as antisense oligonucleotide for selectively
CC   inhibiting expression of the cofactor gene to determine physiological
CC   responses. The protein is useful for screening drugs for agonist and
CC   antagonist activity, and therefore, for screening drugs useful in
CC   regulating physiological responses associated with the cofactors such as
CC   neurosystemic diseases, immunological indications, hormonal dysfunction,
CC   metabolic disorders, immunological responses associated with the
CC   antibodies for detection of the proteins. The polynucleotide can be used
CC   to design primers for a polymerase chain reaction and are also used to
CC   model the three-dimensional structure of the protein. This sequence
CC   represents the complement of the CF6 gene sense coding region.
XX
SQ   Sequence 420 BP; 103 A; 98 C; 64 G; 155 T; 0 U; 0 Other;
```

US-09-701-618A-2 (1-141) x ADI28091 (1-420)

Alignment Scores:

Pred. No.:			
Score:	2.08e-73		
Percent Similarity:	95.71%	Length:	420
Percent Local Similarity:	93.57%	Matches:	131
Best Local Similarity:		Conservative:	3
Query Match:	93.29%	Mismatches:	6
DB:	6	Indels:	0
		Gaps:	0

```
  1 MetAlaGlyGluGluIleLeuAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla 20
    ||||||||||||||||||||||||||||::|||||||||||||||||||||||
420 ATGGCAGGTGAAGAAATTAATGAGGACTATCCAGTAGAAATTCACGAATATTTGTCAGCA 361

 21 PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
    |||
360 TTTGCGAATTCCATTGATGGTGTGGACATGCTGAAGACAATGATGTCTGTTTCGTAGA 301

 41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAlaSerValSerAla 60
    |||||||||||||||||||||||||||||||||||||
300 AATGAGTTGTTGCAGAAGTTGGACCCACTGGAACAAGCAAAAGTGGCATCCGTTTCTGCA 241

 61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
    |||||||||||||||||||||||||||||||||||||
240 TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAGGAGTGAATCCTAAGGA 181

 81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
    ||||||||||||||||||||||||||||||||||||
180 CATCCAGTAAAGCAGGAATTGGAAAGAATCAGAGTATATATGAACAGAGTCAAGGAAATA 100

101 ThrAspLysLysLysAlaLeuGlyLysLysLeuAspAspAlaSerArgPheValLysAsn 120
    ||||||||||||||||||||||::|||||||::|||
180 ACAGACAAGAAAAAAGGCTGGCAAGCTCGGACAGAGGTGCAGCTTCAAGATTTGTAAGAAAT 120

121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAsnLysGlyLysSerLys 140
    |||||||||||||||||||||||||
 61 GCCCTCTGGGAACCAAAATCCAAAGTTGCCCATAAAGGAAAAAGTAAA
```

Db　　　60 GCCCTCTGGGAACCAAAAACCGAAAAATACATCCAAAAGTTGCCCATAAAGGAAAAAAGTAAA 1

RESULT 11

XX
ID　　ADJ92841
XX
AC　　ADJ92841;
XX
ID　　ADJ92841 standard; DNA; 426 BP.
XX
DT　　06-MAY-2004　(first entry)
XX
DE　　Human co-repressor SUN-COR DNA.
XX
KW　　Nuclear receptor; therapy; diabetes; hyperlipidaemia; obesity;
KW　　atherosclerosis; human; co-repressor; gene; ds.
XX
OS　　Homo sapiens.
XX
PN　　US2003228607-A1.
XX
PD　　11-DEC-2003.
XX
PF　　14-APR-2003; 2003US-00414692.
XX
PR　　15-APR-2002; 2002US-0372650P.
XX
PA　　(WAGN/) WAGNER B L.
PA　　(SCHU/) SCHULMAN I G.
XX
PI　　Wagner BL, Schulman IG;
XX
DR　　WPI; 2004-167207/16.
XX
PT　　Identifying compounds that bind to nuclear receptor and exhibit cell type
PT　　specific actions, and useful for treating hyperlipidemia, obesity and
PT　　diabetes.
XX
PS　　Disclosure; SEQ ID NO 42; 99pp; English.
XX
CC　　The invention relates to screening methods for identifying compounds that
CC　　bind to nuclear receptor and exhibit cell type specific actions. The
CC　　invention relates to modulators having an improved therapeutic profile.
CC　　The method is useful for identifying compounds that bind to a nuclear
CC　　receptor and exhibit cell type specific actions. It is also useful for
CC　　identifying modulators of nuclear receptors that are useful in treating
CC　　diseases e.g., diabetes, hyperlipidaemia, obesity, atherosclerosis, etc.
CC　　The present sequence is human co-repressor DNA used to illustrate the
CC　　method of the invention.
XX
SQ　　Sequence 426 BP; 150 A; 76 C; 107 G; 93 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:　　　　　　　　　　　8.74e-72
Percent Similarity:　　　　96.43%
Best Local Similarity:　　90.71%
Query Match:　　　　　　　　91.47%
DB:　　　　　　　　　　　　　12

Length:　　　　426
Matches:　　　　127
Mismatches:　　　5
Conservative:　8
Indels:　　　　　0
Gaps:　　　　　　0

US-09-701-618A-2 (1-141) x ADJ92841 (1-426)

Qy　　　1 MetAlaGlyGluGlnIleAsnGluAspThrProValGluIleHisGluTyrLeuSerAla 20
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　　1 ATGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAATTCACGAGTCTTTAACAGCC 60

Qy　　21 PheGluAsnSerIleGlyValAlaValAspGluMetLeuLysThrMetMetSerValSerArg 40
　　　　　|||:|||||||||||||::|||||||||||:|||||||||||||||::|||||||||||||
Db　　61 CTGGAGAGCTCCCCGGGTCGTGCTGTGGATGACATGCTGAAGACCATGATGTCGGTTTCTAGA 120

Qy　　41 AsnGluLeuLeuGlnLysValAspProLeuGlnLnAlaLysValSerAla 60
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　121 AATGAGTTGTTGCAGAAGTTGGACCCATTGGAACAAAGGCAAAAGGTGGATTTTAGTTTCTGCA 180

Qy　　61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyAlaAsnProLysGlu 80
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db　　181 TACACCTTAAATTCAATGTTTTGGGTTTATTTGGGCAACTCAAGGAGGTTAATCCAAAGAG 240
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　81 HisProValLysGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　241 CATCCAGTGAAGCAGGAACTGGAAAGAATCAGAGTGTACATGAACAGAGTTAAAGAAATA 300

RESULT 12

XX
ID　　AAZ43928
XX
AC　　AAZ43928;
XX
ID　　AAZ43928 standard; cDNA; 1038 BP.
XX
DT　　17-MAR-2000　(first entry)
XX
DE　　Murine C1D cDNA.
XX
KW　　C1D; murine; apoptosis; tumour; gene therapy; treatment; ss.
XX
OS　　Mus sp.
XX
FH　　Key　　　　　　　Location/Qualifiers
FT　　CDS　　　　　　　76..501
FT　　　　　　　　　　/*tag= a
FT　　　　　　　　　　/product= "C1D"
XX
PN　　DE19824811-A1.
XX
PD　　09-DEC-1999.
XX
PF　　03-JUN-1998; 98DE-01024811.
XX
PR　　03-JUN-1998; 98DE-01024811.
XX
PA　　(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI　　Rothbarth K, Stammer H, Werner D;
XX
DR　　WPI; 2000-063506/06.
DR　　P-PSDB; AAY51025.
XX
PT　　Inducing apoptosis by overexpressing the C1D gene, particularly for
PT　　treating tumors.
XX
PS　　Claim 3; Fig 2; 10pp; German.
XX
CC　　This invention describes a novel method for inducing apoptosis which
CC　　comprises overexpressing the C1D gene (I). The method is particularly
CC　　used to treat tumors and can also be used in gene therapy. The method has
CC　　no side effects on normal cells (contrast known methods of inducing
CC　　apoptosis such as cytotoxins and radiation), and may be effective on
CC　　cells resistant to conventional treatments. Overexpression of (I) is
CC　　sufficient itself to induce apoptosis but the effect may be increased
CC　　when used in combination with other anti-tumor methods. When cells
CC　　transfected with (I) undergo apoptosis, they release factors that kill
CC　　neighboring, non-transfected cells (bystander effect). This sequence
CC　　encodes the murine C1D protein described in the method of the invention
XX
SQ　　Sequence 1038 BP; 323 A; 163 C; 219 G; 333 T; 0 U; 0 Other;

Db
Qy　　101 ThrAspLysLysGlyAlaGlyValLysLeuAspArgGlyValAlaAlaSerArgPheValLysAsn 120
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　301 ACAGACAAGAAGGAGGGTGCCAAGCTGGACAGAGGGGTGCTGCTTCAGAGCTTCAGAGAAC 360

Qy　　121 AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaAlaAsnLysGlyLysSerLys 140
　　　　　|||||||||||||||||||||||||||||||||||:|||:||||||||||||||||
Db　　361 GCAACTCTGGGAACCCAAAGCAAAAAACGACACCAAAGGGCTAATAAAAGGGAAAAGCAAA 420

Alignment Scores:
Pred. No.:
Score:　　　　　　　　　　　2.88e-70
Percent Similarity:　　　　95.00%
Best Local Similarity:　　90.00%

Length:　　　　1038
Matches:　　　　126
Conservative:　7
Mismatches:　　7

Query Match:

```
Qy:
DB:

Query Match:                    90.35%        Indels:    0
                                              Gaps:      0

US-09-701-618A-2 (1-141) x AAZ43928 (1-1038)
```

```
Qy      1  MetAlaGlyGluGluIleAsnGluAsnPtyrProValGluIleHisGluTyrLeuSerAla  20
Db     76  ATGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAATTCACGAGTCTTTAACAGCC  135

Qy     21  PheGluAsnSerIleGlyAlaValAspGluMetLeuLysThrMetSerValSerArg  40
           |::::||||||||||||||||||||||||||::|||::||||||||||||||||||||
Db    136  CTGGAGAGCTCCCTGGGTGCTGTTGACGAGATGCTGAAGACCATGTGGGCTGTTTCTAGA  195

Qy     41  AsnGluLeuLeuGlnLysLysValSerLeuAspProLeuAlaLysValAlaSerAla  60
Db    196  AACGAGTTGTTGCAGAAGAAGGTGAGCCTGGACCCATTGGAACAAGGCAAAAGGTGGATTAGTTTCTGCA  255

Qy     61  TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAlaAsnProLysGlu  80
Db    256  TACACCTTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAGGAGTTAATCCCAAAGAG  315

Qy     81  HisProValLysGlnLeuGlnLeuArgIleArgValTyrMetAsnArgValLysGlyIle  100
Db    316  CATCCAGTGAAGCAGCAGAACTGGAGAAGAATCAGAGTGTACATGAACAGAGTAAAGGGAATA  375

Qy    101  ThrAspLysLysLysAlaGlyLysLysLeuAspArgArgGlyLysAlaSerArgPheValLysAsn  120
Db    376  ACAGACAAGAAGAAGGCTCCAAGCTGGACGAGAGGGTGCTTCGAGATTTGTCAGAAG  435

Qy    121  AlaLeuTrpGluProLysSerLysAsnAlaSerLysValAlaLeuAsnLysSerGlyLysSerLys  140
Db    436  GCACTCTGGGAACCCAAACGAAAAGCACACCAAAAGTGGCTTAATAAAGGGAAAAAGCAAA  495
```

Db;

RESULT 13

ID AAL43975 standard; cDNA; 420 BP.

XX

AC AAL43975;

XX

DT 27-SEP-2002 (first entry)

XX

DE Mammalian nuclear receptor cofactor CF8 coding sequence.

XX

KW Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;
KW metabolism regulation; cell homeostasis; cell proliferation;
KW differentiation; pathological cellular aberration;
KW cellular defence mechanism.

OS Mammalia.

XX

FH Key   Location/Qualifiers

FT CDS   1..420
FT   /*tag= a
FT   /partial
FT   /product= "Mammalian CF8 protein"
FT   /note= "No start or stop codon is given"

XX

PN WO200242322-A2.

XX

PD 30-MAY-2002.

XX

PF 21-NOV-2001; 2001WO-EP013548.

XX

PR 21-NOV-2000; 2000EP-00125524.

XX

PA (LION-) LION BIOSCIENCE AG.

XX

PI Jackson D, Casari G, Suckow J;

XX

DR WPI; 2002-566559/60.
DR P-PSDB; AA015406.

XX

PT Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators

---

PT useful for inhibiting cellular function of cofactor and for treating
PT metabolic disorders, immunological indications and hormonal dysfunctions.

XX

PS Claim 1; Fig 1; 68pp; English.

XX

CC The invention comprises the amino acid and coding sequences of two
CC mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC CF8 protein sequences of the invention are useful for screening agents
CC that are capable of inhibiting the cellular function of cofactor CF7
CC and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC involved in cellular functions, such as: regulation of metabolism and
CC cell homeostasis, cell proliferation and differentiation, pathological
CC cellular aberrations, or cellular defence mechanisms. The present cDNA
CC sequence encodes the mammalian nuclear receptor cofactor CF8 protein

SQ Sequence 420 BP; 153 A; 62 C; 100 G; 105 T; 0 U; 0 Other;

Alignment Scores:

```
Pred. No.:                          1.5e-70
Score:                              644.00
Percent Similarity:                 92.86%      Length:           420
Best Local Similarity:              91.43%      Matches:          128
Query Match:                        90.07%      Conservative:     10
DB:                                 6           Mismatches:       2
                                                Indels:           0
                                                Gaps:             0
```

US-09-701-618A-2 (1-141) x AAL43975 (1-420)

```
Qy      2  AlaGlyGluGluIleAsnGluAsnPtyrProValGluIleHisGluTyrLeuSerAlaPhe  21
           |||||||||||||||||||||||||:||||||||||||||||||||||||||||:||||
Db      1  GCAGGTGAAGAAATTAATGAGAACTATCCAGTAGAAATTCACGATTATTTGTCAGCATTT  60

Qy     22  GluAsnSerIleGlyAlaValAlaAspGluMetLeuLysThrMetSerValSerArgAsn  41
           |||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db     61  GCGAATTCCATTGATGCTGTGGATGAGATGCTGAAGAACATGAGTGTCTGTTTCTAGAAAT  120

Qy     42  GluLeuLeuGlnLysLysValSerLeuAspProLeuAlaLysValAlaSerAlaTyr  61
Db    121  GAGTTGTTGCAGAAGAAGGTGGACCCACTTGAACAAGCCAAAAGTGGATTTCTGCATAC  180

Qy     62  ThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAlaAsnProLysGluHis  81
Db    181  ATATTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAGGAGTGAATCCTAAGGAACAT  240

Qy     82  ProValLysGlnLeuGlnLeuArgIleArgValTyrMetAsnArgValLysGlyIleThr  101
Db    241  TCAGTAAAGCAGCAATTGCAGCTGAGAATCAGAGTATATGAACAGAGTCAAGGAATAACA  300

Qy    102  AspLysLysLysAlaGlyLysLysLeuAspArgArgGlyLysAlaSerArgPheValLysAsnAla  121
Db    301  GACAAGAAAAAAGCTGGCAAGAAGCTGGACAGAAGGTGCAAGGCTTCAAGATTTGTAAGAATGCC  360

Qy    122  LeuTrpGluProLysSerLysAsnAlaSerLysValAlaLeuAsnLysSerGlyLysSer  141
Db    361  CTCTGGGAACCAAAACCGAAAAACGATCCATCCAAAGTTGCCTTATAAAGGGAAAAAGT  420
```

Db;

RESULT 14

ID AAL43976/c standard; cDNA; 420 BP.

XX

AC AAL43976;

XX

DT 27-SEP-2002 (first entry)

XX

DE Mammalian nuclear receptor cofactor CF8 reverse complement sequence.

XX

KW Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;
KW metabolism regulation; cell homeostasis; cell proliferation;
KW differentiation; pathological cellular aberration;
KW cellular defence mechanism.

OS Mammalia.

XX

PN WO200242322-A2.

XX
PD   30-MAY-2002.
XX
PF   21-NOV-2001; 2001WO-EP013548.
XX
PR   21-NOV-2000; 2000BP-00125524.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Jackson D, Casari G, Suckow J;
XX
DR   WPI; 2002-566559/60.
XX
OS
XX
PS   Claim 1; Fig 2; 68pp; English.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.
XX
CC   The invention comprises the amino acid and coding sequences of two
CC   mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC   CF8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CF7
CC   and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC   involved in cellular functions, such as: regulation of metabolism and
CC   cell homeostasis, cell proliferation and differentiation, pathological
CC   cellular aberrations, or cellular defence mechanisms. The present cDNA
CC   sequence represents the reverse complement of the mammalian nuclear
CC   receptor cofactor CF8 gene sequence
XX
SQ   Sequence 420 BP; 105 A; 62 C; 153 G; 153 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1.5e-70
Score: 644.00   Length: 420
Percent Similarity: 92.86%   Matches: 128
Best Local Similarity: 91.43%   Conservative: 2
Query Match: 90.07%   Mismatches: 10
                                        Indels: 0
DB: 6   Gaps: 0

US-09-701-618A-2 (1-141) x AAL43976 (1-420)

Qy   2 AlaGlyGluGluGlnIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAlaPhe   21
         |||||
Db 420 GCAGCTGAGAGAAATTAATGAGGACTATACCAGTAGAAATTCACGATTATTTGTCACCATTT 361

Qy   22 GluAsnSerIleGlyAlaValAlaValAspGluMetLeuLysThrMetSerValSerArgAsn   41
         ||||||||
Db 360 GCCGAATTCCATTGATGCTGTGGATGAGATGCTGAAGAACATGATAGTCTGTTTCTTGAAAT 301

Qy   42 GluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAlaTyr   61
         ||||||||
Db 300 GAGTTGTTGCAGAAGTTGGAACCCACTTGAACAAGCAAAAGTGGATTTGGTTTCGCATAC 241

Qy   62 ThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGluHis   81
         ||||||||
Db 240 ATATTAAAATTCAAAGTTTTGGGCTTATTTGGCACTCAAGGAGTGAATCCTAAGGAACAT 181

Qy   82 ProValLysGlnGlnLeuGluArgIleArgValTyrMetAsnArgValLysGluIleThr 101
         ||||||||
Db 180 TCAGTAAAGCAGCAGAATTGGAAAGAATCAGAGTATATATGAACACAGAGTCAAGGAAATAGCA 121

Qy 102 AspIleLysLysAlaGlyLysLeuAsnArgGlyLysAlaAlaSerArgPheValLysAsnAla 121
         ||||||||
Db 120 GACAGAAAAAAGGCTGGCAAGCTCGACACAGAGGTGCAGCTTCAAGATTTGTAAGAAATGCC   61

Qy 122 LeuTrpGluProLysGlySerAsnAlaSerLysValAlaAsnLysGlyLysSerLysSer 141
         ||||||||
Db   60 CTCTGGGAACCAAAAGGCATCCAAAGTTGCCTATAAAGGAAAAAGTAAAAGT   1

XX
AC   AAC01237;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 1235.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG01231.
XX
PR   26-FEB-1999; 99US-0122487P.
XX
PF   21-FEB-2000; 2000BP-00200610.
XX
PD   06-SEP-2000.
XX
PS   Claim 1; SEQ ID NO 1235; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors
XX
SQ   Sequence 380 BP; 119 A; 59 C; 95 G; 106 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:   4.39e-55
Score: 519.00   Length: 380
Percent Similarity: 100.00%   Matches: 102
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 72.59%   Mismatches: 0
                                        Indels: 0
DB: 3   Gaps: 0

US-09-701-618A-2 (1-141) x AAC01237 (1-380)

Qy   1 MetAlaGlyGluGluGlnIleAsnGluAspTyrProValGluIleHisGluTyrLeuSerAla   20
         ||||||||
Db   1 ATGGCAGGTGAAGAATTAATGAAGACTATCCAGTAGAAATTCACGATTATTTGTCAGCG 134

Qy   21 PheGluAsnSerIleGlyAlaValAlaValAspGluMetLeuLysThrMetSerValSerArg   40
         ||||||||
Db 135 TTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGAGTGTCTGTTTCTAGA 194

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysValAspLeuValSerAla   60
         ||||||||
Db 195 AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA 254

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu   80

ID   AAC01237 standard; cDNA; 380 BP.
XX
AC   AAC01237
RESULT 15

```
Db  255  TACACATTAAATTCAAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAAATCCTAAGGAA  314
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81       HisProValLysGlnLeuGluArgIleArgValTyrMetAsnArgValLysGluIle  100
Db  315  CATCCAGTAAAACAGGAATTGGAAAAGAAATCAGAGTTATATATGAACAGAGTCAAGGAAATA  374
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  101                                                          ThrAsp  102
Db  375  ACAGAC  380
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 4, 2005, 07:11:31

Title: US-09-701-618A-3
Perfect score: 1040
Sequence: 1 cagaagccgtcatcggcgt..........tcattaacttctgatggaaaa 1040
Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0
Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
                 1: gb_est1:*
                 2: gb_est2:*
                 3: gb_htc:*
                 4: gb_est3:*
                 5: gb_est4:*
                 6: gb_est5:*
                 7: gb_est6:*
                 8: gb_gss1:*
                 9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	958	92.1	2674	3	AK028702	AK028702 Mus muscu
2	952	91.5	2958	3	AK035169	AK035169 Mus muscu
3	883.8	85.0	1002	2	BB609670	BB609670
4	805.4	77.4	876	7	CB588136	CB588136 AGENCOURT
5	805.2	77.4	864	7	CF551758	CF551758 AGENCOURT
6	768.2	73.9	871	5	BU151822	BU151822 AGENCOURT
7	767.4	73.8	820	5	BU525697	BU525697 AGENCOURT
8	694.4	66.7	733	5	CB317510	CB317510 AGENCOURT
9	692	66.5	731	6	CB574390	CB574390 AGENCOURT
10	686	66.0	715	6	CB598466	CB598466 AGENCOURT
c 11	678.8	65.3	781	1	AI875855	AI875855 uj67b10.x
12	678.2	65.2	685	2	BB612968	BB612968
13	661.6	63.6	690	5	BQ746434	BQ746434 UI-M-ER0-
14	654.2	62.9	671	6	CD352734	CD352734 UI-M-GL0-
15	651.6	62.7	772	5	CB320669	CB320669 AGENCOURT
16	650.2	62.3	657	2	BB627449	BB627449 UI-M-BH3-
17	648.4	62.3	678	5	BM936857	BM936857 AGENCOURT
18	643.6	61.9	836	4	BI694343	BI694343 60346907
19	640.4	61.6	667	7	CK794480	CK794480 60346907
20	636.2	61.2	642	6	BQ564508	BQ564508 gi18a09.y
21	634	61.0	668	5	CD773363	CD773363 AGENCOURT
22	632.6	60.8	737	4	BI697211	BI697211 60334715
23	631.6	60.7	753	6	CB571517	CB571517 AGENCOURT
24	615	59.1	712	6	CB950526	CB950526 AGENCOURT
25	612.2	58.9	617	4	BG081538	BG081538 H3066D01-
26	579	55.7	640	6	CD766200	CD766200 AGENCOURT
27	562.6	54.1	648	2	AW611366	AW611366 un63h12.x
28	561.2	54.0	661	2	AW986301	AW986301 uf77f06.x
c 29	557.4	53.6	659	6	CD775407	CD775407 UI-M-AQ0-
30	557.2	53.6	562	4	BM125918	BM125918 id77a03.y
31	552.2	53.1	657	1	AV002444	AV002444 AGENCOURT
32	543.4	52.2	750	6	CB572493	CB572493 AGENCOURT
33	542	52.1	761	6	CB953171	CB953171 AGENCOURT
c 34	541.4	52.1	618	2	BB688895	BB688895 uw1a04.x
35	540.2	51.9	766	6	CB574401	CB574401 AGENCOURT
36	535.4	51.5	629	6	CD773175	CD773175 AGENCOURT
37	534.4	51.4	544	2	AW475291	AW475291 mr13h04.r
38	533.2	51.3	549	1	AA153178	AA153178 mr17h04.r
39	528.2	50.8	650	6	CF110457	CF110457 Shiltzomi
40	527.2	50.7	532	4	BM070131	BM070131 id78b03.y
41	527.2	50.7	781	4	BG862190	BG862190 602795770
42	525.2	50.5	534	1	AI786504	AI786504 ufi5g05.y
43	521.2	50.1	526	1	AA476164	AA476164 vh21h09.r
44	519.2	49.9	795	8	AF046558	AF046558 AF046558
45	519.2	49.9	795	9	CG478915	CG478915 OS79319 M

ALIGNMENTS

RESULT 1
AK028702   2674 bp   mRNA   linear   HTC 03-APR-2004
LOCUS   AK028702
DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
            library, clone:4732439908 product:nuclear DNA binding protein, full
            insert sequence.
ACCESSION   AK028702
VERSION   AK028702.1 GI:26324641
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   MEDLINE 20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,K.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   MEDLINE 20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research

```
TITLE       Group Phase I & II Team.
            Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 2674)
AUTHORS     Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,Y., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Miramatsu,M. and Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            Fax:81-45-503-9216)
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://fantom.gsc.riken.jp/
            URL:http://genome.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1..2674
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM DB:4732439J08"
                     /db_xref="taxon:10090"
                     /clone="4732439J08"
                     /tissue_type="skin"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     CDS             58..483
                     /note="unnamed protein product; nuclear DNA binding
                     protein (MGD|MGI:1927354, GB|NM_020558, evidence: BLASTN,
                     98%, match=1021)"
                     /codon_start=1
                     /product="putative"
                     /protein_id="BAC26075.1"
                     /db_xref="GI:26324642"
                     /translation="MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTWMAVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVVLATOGVNPKEHPVKOELERIRVYMRVKEITD
                     KKKAAKLDRGAAASRFVKNALWEPKAKSTPKVANKGKSKH"
     polyA_signal    2661..2666
                     /note="putative"
     polyA_site      2674
                     /note="putative"
ORIGIN
```

```
Query Match             92.1%;  Score 958;  DB 3;  Length 2674;
Best Local Similarity   98.6%;
Matches 1008;  Conservative 0;  Mismatches 10;  Indels 4;  Gaps 4;

Qy         22 ATCATCGTGCGACCTATTTCCCGAGACAGGCGCTCCACGGTAGTATTGAGTTGGTCACAATGG 81
              ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          2 ATCATCGTGCGACCTATTTCCCGAGACAGGCGCTCCACGGTAGTATTGAGTTGGTCACAATGG 61

Qy         82 CAGGTGAAGAAATGAAATGAAGAGTATCCGGAGAAGAATTCACGGAGTCTTAACAGCCCTGG 141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CAGGTGAAGAAATGAAATGAAGAGTATCCGGAGAAGAATTCACGGAGTCTTAACAGCCCTGG 121

Qy        142 AGAGCTCCCTGGGTGCGTGGACGACATGGCTCGAAGACCATGAATGGCGTGTTTCTGAAACG 201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGAGCTCCCTGGGTGCGTGGACGACATGGCTCGAAGACCATGAATGGCGTGTTTCTGAAACG 181

Qy        202 AGTTGTTCCAGAAGTTGGACCCCATTGGAACAAGCAAAGGTGGAATTAGTTGTTCTGCATACA 261
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 AGTTGTTCCAGAAGTTGGACCCCATTGGAACAAGCAAAGGTGGAATTAGTTGTTCTGCATACA 241

Qy        262 CCTTAAATTCAATGTTTTTATGGGTTTATTTGGCAACTCCAAGGAGTTAATCCCAAAGAGCATC 321
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CCTTAAATTCAATGTTTTTATGGGTTTATTTGGCAACTCCAAGGAGTTAATCCCAAAGAGCATC 301

Qy        322 CAGTGAAGCAGGAAACTGGAAAAGAAATCGAGTCTACAGAGAGTTAAAGAAATAACAG 381
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 CAGTGAAGCAGGAAACTGGAAAAGAAATCGAGTCTACAGAGAGTTAAAGAAATAACAG 361

Qy        382 ACAAGAGAGGAAGGCTGCCAAGCTGGACAGAGGTGCTGCTTCGAGATTTGTCAAGGAAGGCAC 441
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ACAAGAGAGGAAGGCTGCCAAGCTGGACAGAGGTGCTGCTTCGAGATTTGTCAAGGAAGGCAC 421

Qy        442 TCTGGGAACCCAAAACGACACCAAAAGGGGAAAAGCAAAACACT 501
              |||||| |||||||||||||||||||||||||||||||||||||
Db        422 TCTGGGAACCCAAAACGACACCAAAAGGGGAAAAGCAAAACACT 481

Qy        502 AATCTTTTGGTTTGTTTGATGTACATGTTTTCAAAAAGTACATCCTTTTTTAATCAGTTTACA 561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AATCTTTTGGTTTGTTTGATGTACATGTTTTCAAAAAGTACATCCTTTTTTAATCAGTTTACA 561

Qy        562 TGTACTTATGTGACCATGTGGTGTGTTTAAATGGAATCCTTTTGGAATTCAATGTATAAAATT 621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGTACTTATGTGACCATGTGGTGTGTTTAAATGGAATCCTTTTGGAATTCAATGTATAAAATT 601

Qy        622 ACACATTACATTTGTGATACTGAATCTTTTTTTGCTGAGAAGAGATTAAGTTGTCTTTTGT 681
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 ACACATTACATTTGTGATACTGAATCTTTTTTTGCTGAGAAGAGATTAAGTTGTCTTTTGT 661

Qy        682 TGATTTCATTAAAGCATCATGAGTGTTTAATATGTTAATGTAAGGGGAAAAGCAAACACT 741
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        662 TGATTTCATTATAAAGCATCATGAATGTGTTTAATATGTTAAGTATTTCATAAGCAGTTG 721

Qy        742 TGAAATCCAAAATGTTCTCTGTGTAAACATTTGTAGTTGTTTGTTAAATGAACAATGATTATGA 801
              ||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        722 TGAAATCCAAAATGTTCTCTGTGTAAACATTTGTAGTTGTTTGTTAAATGAACAATGATTATGA 781

Qy        802 AGTGTGTTGCTTATCTGTAGACCTCGAGGTGTAAGGACACATTTGTTTTCAGTTAATGATGAGAAT 861
              |||||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||||||
Db        782 AGTGTGCTTATCTGTAGA-CTGGAGGTGGTAAGACACATTTGTTTTCAGTTAATGATGAGAAAT 840

Qy        862 ACAGTGACTTAAAATACCCACTCTG-TTTCTGTTCAGTTAGTTCAACATGTTTCGTGGATTT 920
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        841 ACAGTGACTTAAATACCCACTCTGTTTTCTGTTCAGTTAGTTCAACATGTTTCGTGGATTT 900

Qy        921 TTTTTTTTTTTTGAGTAAATT-CTGTCTTGATATTCAAAGT-CAAAATTGAAAACCTTACG 978
              ||||||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||
Db        901 TTTTTTTTTTTTGAGTAAATTGCTGTCTTGATATTCAAAGTTCAAAATTGAAAACCTTACG 960

Qy        979 CTGTACTTTAATTCTTCATGTTCCATTAAAATAAAATGTTCTCATTAACTCTGATGGAA 1038
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGTACTTTAATTCTTCATGTTCCATTAAAATAAAATGTTCTCATTAACTCTGATGGAA 1020

Qy       1039 AA 1040
              ||
Db       1021 AA 1022
```

```
KEYWORDS     HTC; CAP trapper.
SOURCE       Mus musculus
  ORGANISM   Mus musculus (house mouse)
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1
  AUTHORS    Carninci,P. and Hayashizaki,Y.
  TITLE      High-efficiency full-length cDNA cloning
  JOURNAL    Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE    99272253
  PUBMED     10349636
REFERENCE    2
  AUTHORS    Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE      Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL    Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE    20499974
  PUBMED     11042159
REFERENCE    3
  AUTHORS    Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,Y., Ozawa,K., Tanaka,T., Matsuura,C., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE      RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
  JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE    20530913
  PUBMED     11076861
REFERENCE    4
  AUTHORS    The RIKEN Genome Exploration Research Group Phase II Team and the
             FANTOM Consortium.
  TITLE      Functional annotation of a full-length mouse cDNA collection
  JOURNAL    Nature 409, 685-690 (2001)
REFERENCE    5
  AUTHORS    The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
  TITLE      Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL    Nature 420, 563-573 (2002)
REFERENCE    6  (bases 1 to 2958)
  AUTHORS    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hiramatsu,T., Hiraoka,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
             URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT      cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in Riken
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.jp/.
             URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1..2958
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:9430096A21"
                     /db_xref="taxon:10090"
                     /clone="9430096A21"
                     /tissue_type="embryonic body between diaphragm region and
                     neck"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="12 days embryo"
     CDS             52..477
                     /note="unnamed protein product; nuclear DNA binding
                     protein (MGD|MGI:1927354, GB|NM_020558, evidence: BLASTN,
                     98%, match=1021)"
                     /codon_start=1
                     /protein_id="BAC28967.1"
                     /db_xref="GI:26330474"
                     /translation="MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMWAVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITD
                     KKKAAKLDRGAASRFVKNALWEPKAKSTPKVANKGKSKH"
ORIGIN
             Query Match    91.5%;  Score 952;  DB 3;  Length 2958;
             Best Local Similarity  98.6%;  Pred. No. 3.7e-204;
             Matches 1002; Conservative   0; Mismatches 10; Indels  4; Gaps  4;

Qy         28 GTGCGACCTATTTCCCGGAGACAGGCGGTCCACGGGTATTGAGTGGTGGTCACAATGGCAGGTG 87
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GTGCGACCTATTTCCCGGAGACAGGCGGTCCACGGGTATTGAGTGGTGGTCACAATGGCAGGTG 61

Qy         88 AAGAAATGAATGAAGAGATTATCCCGTAGAAAATTCACCGAGTCTCTTAACAGCCCTGGGAGAGCT 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db         62 AAGAAAATGAATGAAGAGATTATCCCGTAGAAAATTCACCGAGTCTCTTAACAGCCCTGGGAGAGCT 121

Qy        148 CCCTGGGGTGCTGTGGACGACATGCTGAGAGACCATGATGGCTGTCTTCTAGAAACGAGTTGT 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CCCTGGGGTGCTGTGGACGACATGCTGAGAGACCATGATGGCTGTCTTCTAGAAACGAGTTGT 181

Qy        208 TGCAGAAGTTGGACCCCATTGGAACAAGCAAAAGGTGGATTTAGTTGTTTCTGCATACACCTTAA 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TGCAGAAGTTGGACCCCATTGGAACAAGCAAAAGGTGGATTTAGTTGTTTCTGCATACACCTTAA 241

Qy        268 ATTCAATGTTTTTGGGGTTTTATTTGGGCAACTCAAGGAGTTAATCCCAAAGAGCATCCAGTGA 327
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ATTCAATGTTTTTGGGGTTTTATTTGGGCAACTCAAGGAGTTAATCCCAAAGAGCATCCAGTGA 301

Qy        328 AGCAGGAACTGGAAGAATCAGAGTCTACATGAACAGAGTTAAAGAGAAATAACAGACAAGA 387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AGCAGGAACTGGAAGAATCAGAGTCTACATGAACAGAGTTAAAGAGAAATAACAGACAAGA 361

Qy        388 AGAAGGCTGCCAAGCTGGACAGAGGTGCTGCTTCAGAGATTGGTCAAGAGGCCACTCTGGG 447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AGAAGGCTGCCAAGCTGGACAGAGGTGCTGCTTCAGAGATTGGTCAAGAGGCCACTCTGGG 421

Qy        448 AACCCAAAGCAAAAAGCACACCAAAAGTGGCTAATAAAGGGAAAAGCAAAACACTAATCTT 507
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 AACCCAAAGCAAAAAGCACACCAAAAGTGGCTAATAAAGGGAAAAGCAAAACACTAATCTT 481

Qy        508 TTGGGTTTTGATGTACATGTTTTCAAAAAGTACATCCTTTTTTAATCAGTTTTACCAATGTAGT 567
              ||||||||| |||||||||||||||||||||||||||||||| |||||||||||||||||||
Db        482 TTGGGTTTTGATGTACATGTTTTCAAAAAGTACATCCTTTTTTAATCAGTTTTACCAATGTAGT 541

Qy        568 TATGTGACCATGTGGTGTGTTTTAAATGGATTCCTTTTGGAATTCATGTATAAATTTACACAT 627
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        542 TATGTGACCATGTGGTGTGTTTTAAATGGATTCCTTTTGGAATTCATGTATAAATTTACACAT 601

Qy        628 TACATTTGTGATACTGAAATCTCTTTTTTTTTGCTGAGAAAGATTAAAGTTGTCTTTGTTGATTT 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 TACATTTGTGATACTGAAATCTCTTTTTTTTTGCTGAGAAAGATTAAAGTTGTCTTTGTTGATTT 661

Qy        688 TCAATATAAAGCATCAATGATGTGTTTAATATATTGTAAGATATATTCTATAAGCAGTGTGTGAAAT 747
              ||||||||||||||||||||||||||||||||||||||
Db        662 ...
```

```
RESULT 3

LOCUS       BB609670
DEFINITION  BB609670 RIKEN full-length enriched, 18 days embryo Mus musculus
            cDNA clone 1110036E10 5', mRNA sequence.
ACCESSION   BB609670
VERSION     BB609670.1  GI:16451385
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1002)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.

FEATURES             Location/Qualifiers
     source          1..1002
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="1110036E10"
                     /clone_lib="RIKEN full-length enriched, 18 days embryo"
                     /note="Site 1: XhoI; Site 2: SstI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer (5'
                     GAGAGAGAGAGCGGCCGCAAACTCGAGTTTTTTTTTTTTTTTTTVN 3'), cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence (5'
                     GAGAGAGAGAGAGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCCC 3']."
                     /lab_host="SOLR"
                     /dev_stage="18 days embryo"

ORIGIN
Query Match  85.0%;  Score 883.8;  DB 2;  Length 1002;
Best Local Similarity  97.6%;
Matches  929;  Conservative  0;  Mismatches  18;  Indels  5;  Gaps  3;

Qy      1 CAGAGAGCCGTGTCATGGCGGTCATCATCGTCATCATCGTGCCACCTATTCCGGAGCAGGCAGGGCGTCACG    60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     29 CAGAGAGCCGTGTCATGGCGGTCATCATCGTGCCGACCTATTCCGGAGCAGCAGGGCGTCACG    88

Qy     61 GTATTGAGTTGGGTCACAATGGCAGGTGGAGAAAATGAATGAAGATTATCCCGTAGAAATTC    120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89 GTATTGAGTTGGGTCACAATGGCAGGTGGAGAAAATGAATGAAGATTATCCCGTAGAAATTC    148

Qy    121 ACGAGTCTTTAACAGCCCCTGGAGAGCTCCCGTCGGGTGTGTGGACGACATGCTGAAGACCA    180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 ACGAGTCTTTAACAGCCCCTGGAGAGCTCCCGTCGGGTGTGTGGACGACATGCTGAAGACCA    208

Qy    181 TGATGGCTGGTTTCTTCTAGAAACGAGTTGTTGACAGCCCATTGGAACAAGCAAAGG    240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    209 TGATGGCTGGTTTCTTCTAGAAACGAGTTGTTGCAGAAGTTGGACCCATTGGAACAAGCAAAGG    268

Qy    241 TGGATTTAGTTTCTGCATACACACCTTAAATTCAAATGTTTTGGGTTTTGGTTTATTTGGCAACTCAAG    300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    269 TGGATTTAGTTTCTGCATACACCTTAAATTCAAATGTTTTGGGTTTTGGTTTATTTGGCAACTCAAG    328

Qy    301 GAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGAACTGAAAGAAATCAGAGTCTACACTGA    360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    329 GAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGAACTGAAAGAAATCAGAGTCTACACATGA    388

Qy    361 ACAGAGTTAAAGAAAATAACAGACAAGAAGAAGGCTGCCAAGCACTGGACAGAGAGGTGTGCTT    420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    389 ACAGAGTTAAAGAAAATAACAGACAAGAAGAAGGCTGCCAAGCACTGGACAGAGAGGTGCTT    448

Qy    421 CGAGATTGTCAAGAAGGCACTCTGGGAACCCAAACGAAAAAAGCACACAAAAGTGGGCTA    480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    449 CGAGATTGTCAAGAAGGCACTCTGGGAACCCAAAGCAAAAAAGCACACAAAAAGTGGGCTA    508

Qy    481 ATAAAGGGAAAAGCAAACACTAACTCTTTGGTTTGATGTACATTGTTTCAAAAAAGTACA    540
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    509 ATAAAGGGAAAAGCAAACACTAACTCTTTGGTTTGATGTACATTGTTTCAAAAAAGTACA    568

Qy    541 TCCTTTTTTAATCAGTTTACAAATGTAGTTTATGTACCATGTGTTGTTTAAATGGATTCCTT    600
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    569 ...
```

```
Db    662 TCATATAAAGCAATCATGATGTGTTTAATATTGTAGATATTCTATAGACAGTTGTGTGAAAT    721
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    748 CCAAATGTTCTCTGTAAACAATTTGTAGTGTTTGAAATGAACAATGATATATTATGAAGTGTG    807
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 CCAAATGTTCTCTGTAAACAATTTGTAGTGTTTGAAATGAACAATGATATATTATGAAGTGTG    781
Qy    808 CTATCTGTAGACCTCCGAGGTGTAGGACATTGTTTGTTCAGTAATGATGACAAATACAGTG    867
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    782 CTATCTGTAGA-CTCGAGGTGTAAGGACATTGTTTGTTCAGTAATGATGACAAATACACC    840
Qy    868 ACTTAAAATACCCACTCTG-TTTCTGTTCAGTTAGTTCACAGTGTTTCGTGATTTTTTTTTT    926
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ACTTAAATACCCACTCTGTTTTGTTCAGTTAGTTCACATGTTTCGTGATTTTTTTTTT    900
Qy    927 TTTTTTGAGTAATT-CTGTCTTGATATTCAAAGT-CAAAATGAAACCTTAAGGCTCGTAC    984
            ||||||||||| || |||||||||||||| |||| | ||||||||||||||||||||
Db    901 TTTTTTGAGTAATTGCTGTCTTGATATTCAAAGTACAAAAACCTTAAGGCTGTAC    960
Qy    985 TTTAATTCTTCATTTAAAATAAAATGTTCTCATTAACTCTGATGGAAA    1040
            |||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTTAATTCTTCATGTTCCATTTAAAAATAAAATGTTCTCATTAACTCTGATGGAAA    1016
```

RESULT 4
LOCUS   CB588136

DEFINITION CB588136   876 bp   mRNA   linear   EST 03-APR-2003
            AGENCORT 13039966 NIH_MGC_136 Mus musculus cDNA clone
            IMAGE:30293578 5', mRNA sequence.
ACCESSION   CB588136
VERSION   CB588136.1 GI:29505992
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 876)
  AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM339 row: e column: 11
            High quality sequence start: 7
            High quality sequence stop: 704.
FEATURES   Location/Qualifiers
     source   1..876
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3029578"
                     /tissue_type="embryonic limb, maxilla and mandible"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_136"
                     /note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
                     Normalized, full-length enriched library from pooI of
                     mouse embronic limb, maxilla and mandible, embryonic day
                     17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
                     equivalents from respective days). Cloned directionally,
                     oligo-dT primed (5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCT(T)15-3'.
                     Size selected for the >5kb fragments, average insert size
                     1.2 kb. Normalization to Cot 7.5. Tissue contributed by
                     David Rowe; library constructed by ResGen, Invitrogen
                     Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match   77.4%; Score 805.4; DB 6; Length 876;
Best Local Similarity 96.8%; Pred. No. 4.1e-171;
Matches 843; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

Qy   16 GGCGTCATCATCGTGCGAGAAATGA-ATGAAGATTATCCGGAGAAATTCACGAGTCTTAAC 75
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   6 GTCGTCATCATCGTGCGAGAAATGA-ATGAAGATTATCCGGAGAAATTCACGAGTCTTAAC 65

Qy   76 CAATGGCAGGTGAAGAAATGA-ATGAAGATTATCCGGAGAAATTCACGAGTCTTAACA 134
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   66 CAATGGCAGGTGAAGAAATGA-ATGAAGATTATCCCGTAGAAATTCATGAGTCTTTAAACA 125

Qy   135 GCCCTGGAGAGCTCCCTGGGTGCTGCTGTCCTGGACGACACATGCTGAAGACCATGCTGTT 194
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   126 GCCCTGAGAGCTCCCTGGGTGCTGCTGTCCTGGACGACACATGTTGAAGACCATGATGCCT 185

Qy   195 AGAAACGAGTTGTTGCAGAAGTGGGACCCATTGGGAACAAGCAAAGGT-GGATTAGTTTC 253
              ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   186 AGAAACTAGTTGTTGCAGAAGTGGGACCCATTGGGAACAAGCAAAGGTGGGATTAGTTTC 245

Qy   254 TGCATACACCTTAAATTCAAATGTTTTATTGGGTTTATTGGTTGGCAACTCAAGGAGTTAA 313
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   246 TGCATACACCTTAAATTCAAATGTTTTATTGGGTTTATTGGTTGGCAACTCAAGGAGTTAA 305

Qy   314 AGAGCATCCAGTGAAGCCAGGAACTCAGAGAAGAATCAGAGTCTTACACATGAACAGAGTT 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   306 AGAGCATCCAGTGAAGCCAGGAACTCAGAGAAGAATCAGAGTCTTACACATGAACAGAGTT 365

Qy   374 AATAACAGACAAAGAAGAAGGCTGCCAAGCTGGACACGAGAGTGGCCTGCTTCGAGATTTGTCAA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   366 AATAACAGACAAAGAAGAAGGCTGCCAAGCTGGACACGAGAGTGGCCTGCTTCGAGATTTGTCAA 425

Qy   434 GAAGGCACTCTGGGAACCCAAACAGAAAAAGACACCACCAAAAGTGGCTGGCTTAATAAAAAC 493
              |||| |||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   426 GAACGGCACTCTGGGAACCCAAACAGCAAAACGTGGCTAATAAAGCGGAAAG 485

Qy   494 CAAACACTAATCTTTAATCTTTTGGTTTTTGATGTACATGTTTTCAAAAAGTACATCCTTTTTTAATCA 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   486 CAAACACTAATCTTTAATCTTTTGGTTTTTGATGTACATGTTTTCAAAAAGTACATCCTTTTTTAATCA 545

Qy   554 GTTTACAAATGTAGTTATGTGACCATGTGGGTGTTTAAATGAATTCCTTTTGGAAATTCATGT 613
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   546 GTTTACAAATGTAGTTATGTGACCATGTGGGTGTTTAAATGAATTCCTTTTGGAAATTCATGT 605

Qy   614 ATAAAATTTACACCATTACAATTTGTGACTACATTTGGTAGTTGTTGTTTCAGTTAATGA 673
              |||||||||||||||||| ||||||||||| |||||||||| ||||||||
Db   606 TTAAAATTTACACATTACATTTGTGATACTGAAATCTTTTTTGTTTTTGCTGAGAAAGATTAAGT 665

Qy   674 GTCTTTGTTGATTTTCATATAAAAGCATCATGATGTGTTTTTAATATTGTAATGATAATTCTATA 733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   666 GTCTTTGTTGATTTTCATATAAAAGCATCATGATGTGTTTTTAATATTGTAATGATAATTCTATA 725

Qy   734 AGCAGTTGTGAAAATCCAAAAATGTTCCTCGTGAAACATTTGTAGTGTTTGAAATGAACAATGA 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   726 AGCAGTTGTGAAAATCCAAAAATGTTCCTCGTGAAACATTTGTAGTGTTTGAAATGAACAATGA 785

Qy   794 TATTATGAAGTGTGCTATCGTGAGGTGTAAGGACATTGTAAGGTGTAAGGACTTTATGTTTT 853
              |||| ||||||||||||||| ||||| |||| ||||||||| |||||
Db   786 TATTATGAAGTGTGCTATCGTGAAAACTGGAAGGTGTAAGGACTTTTTGTTTTTTCCATAATGA 845

Qy   854 TGAGAAATACAGTGACTTAAATAACCCACACTCT 884
              ||||||||||||||||||||||||||||||||||
Db   846 TGAAAAATACAGTGACCTTAAATAACCCACACT 876

RESULT 5
LOCUS   CF551758

DEFINITION CF551758   864 bp   mRNA   linear   EST 25-SEP-2003
            AGENCORT 15587320 NIH_MGC_222 Mus musculus cDNA clone
            IMAGE:30521123 5', mRNA sequence.

Left column

```
ACCESSION   CF551758
VERSION     CF551758.1  GI:34888592
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 864)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Naryan Bhat
            cDNA Library Preparation: Express Genomics
            cDNA Library Arrayed by: Agencourt Bioscience Corporation
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM595  row: f  column: 12
            High quality sequence stop: 769.
FEATURES             Location/Qualifiers
     source          1..864
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30521123"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_222"
                     /note="Organ: placenta; Vector: pExpress-1; Site 1: EcoRV;
                     Site 2: NotI; RNA obtained from three placentas from
                     female C57/Bl6 mouse at 16 days pregnancy. Tissues were
                     snap-frozen and kept at -80C for two days before RNA
                     extraction and purification (Tri-reagent method). cDNA was
                     primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. Size-selection >1 kb
                     resulted in an average insert size of 1.5 kb. Library is
                     not amplified. (Normalized version of this library is
                     NIH_MGC_203)."
ORIGIN
```

Query Match

```
Best Local Similarity  77.4%;
Matches  830;  Conservative  0;  Mismatches  23;  Indels  1;  Gaps  1;
Score 805.2;  DB 7;  Length 864;
```

```
Qy   51 GGGCGTCCACGGTATTCAGTTGGTCACTACCAAATGGCAGCTGGAGAAGAAAATGATGAAGAAGATTATCCC 110
        |||||||||||||||||||||||||||| || |||||||||||||||||||||||||||||||||||||||
Db    1 GGCGTCCACGGTATTCAGTTGGTCACAATGGCAGCTGGAGAAGAAAATGATGAAGAAGATTATCCC       60

Qy  111 GTAGAAAATTCACGAGTTCTTTAAACAGCCCTGGGAGGAGCTCCCTGGGTGTCTGTGGACGACATG       170
        |||||||||||||||||||| | |||||||||||||||||||||||||||||||||||||||||||
Db   61 GTAGAAATTCACGAGTCTTTAACAGCCCTGGGAGGAGCTCCCTGGGTGTCTGTGGACGACATG        120

Qy  171 CTGAAGACCATGATGAGGCTGCTTTCTGTTCTTAGAAACGAGTTGTTCGACCCATTGGAA          230
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CTGAAGACCATGATGAGGCTGTTTCTGTTCTTAGAAACGAGTTGTTCGACCCATTGGAA           180

Qy  231 CAAGCAAAGGTGGATGATTTAGTTCTGCATACACCTTAAATTGGGTTTTATTTG               290
        |||||||||||||||| | |||||||||||||||||||||||||| ||||||||
Db  181 CAAGCAAAGGTGGATTTAGTTCTGCATACACCTTAAATTGGGTTTTATTTG                   240

Qy  291 GCAACTCAAGGAGTTAATCCAAAGAGACATCCAGTGAAGCAGGACTGGAGGAATCAGA            350
        |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  241 GCAACTCAAGGAGTTAATCCAAAGAGGCATCCAGTGAAGCAGGACTGGAGGAATCAGA            300

Qy  351 GTCTACATGAACACAGAGAGTTAAAGAAAATAAACAGAGAAAGAAGAAGGCTGCCAAGCTGGACAGA   410
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GTCTACATGAACACAGAGAGTTAAAGAAAATAAACAGAGAAAGAAGAAGGCTGCCAAGCTGGACAGA   360
```

Right column

```
Db  361 GGTGCTGCTTCGAGATTTGTCAGAAGGCAACTCTGGGGAACCCAAAACGAAAAAGCACCA          420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  411 GGTGCTGCTTCGAGATTTGTCAGAAGGCAACTCTGGGGAACCCAAAACGAAAAAGCACCA          470

Db  421 AAAGTGGCTAATAAAGGGAAAAGCAAAACACTAATCTTTTGGTTTTGATGGTACATGTTTTC       480
        ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  471 AAAGTGGCTAATAAAGGGAAAAGCAAAACACTAATCTTTTGGTTTTGATGGTACATGTTTTC       530

Db  481 AAAAAGTACATCCTTTTTAATCAGTTTACAATGTTTATGTTATGTGACCATGTTGGTGTTTAAA     540
        |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Qy  531 AAAAAGTACATCCTTTTTAATCAGTTTACAATGTTTATGTTATGTGACCATGTTGGTGTTTAAA     590

Db  541 TGGATTCCCTTTTGGAATTCATGTATAAAATTTACACATTACACATTTGTGTGATCTTT         600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  591 TGGATTCCTTTTGGAATTCATGTATAAAATTTACACATTACACATTTGTGATCTTT          650

Db  601 TTTTTGCTGAGAAAGAATTAAGTTGTCTTTGTTGATTTTCATATAAAGCATCATCATGATGTCT     660
        |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Qy  651 TTTTTGCTGAGAAAGAATTAAGTTGTCTTTGTTGATTTTCATATAAAGCATCATCATGATGTCT     710

Db  661 TTAATAATTGTAAGATATTCTATAAGCAGTTGTGAAATCCAAAAGTTCTCTGTAAACATTT       720
        |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  711 TTAATAATTGTAAGATATTCTATAAGCAGTTGTGAAATCCAAAAGTTCTCTGTAAACATTT       770

Db  721 GTAGTGTTTTGAAATGAACAAATGATATTATGAAGTGTGCTATATCGTGATTTGTAGA-CTGGAGGTGTA   779
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Qy  771 GTAGTGTTTTGAAATGAACAAATGATATTATGAAGTGTGCTATATCGTGATTTGTAGACTGGAGGTGTA    830

Db  780 AGGACATTTGTTTTCAGTAATGAATGAGAAATACCCACTTCTGTTTCT          890
        |||||||||||||||||||||||||||||||||||||||| |||||||
Qy  831 AGGACATTTGTTTTCAGTAATGAATGAGAAATACCCACTTCTGTTTCT          839

Db  840 TTCTGTTCAGTTTA          853
        ||| || |||||||
Qy  891 GTTCAGTTAGTTCA          904
```

```
RESULT 6
BU151822
LOCUS       BU151822     871 bp     mRNA     linear     EST 03-SEP-2002
DEFINITION  AGENCOURT_8754213 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332444
            5', mRNA sequence.
ACCESSION   BU151822
VERSION     BU151822.1  GI:22665354
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 871)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
            cDNA Library Preparation: ResGen, Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13790  row: a  column: 21
            High quality sequence stop: 615.
FEATURES             Location/Qualifiers
     source          1..871
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6332444"
```

LOCUS   BU525697   820 bp   mRNA   linear   EST 13-SEP-2002
DEFINITION AGENCOURT_10097122 NIH_MGC_144 Mus musculus cDNA clone
            IMAGE:6534234 5', mRNA sequence.
ACCESSION   BU525697
VERSION   BU525697.1 GI:22836150
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 820)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
            cDNA Library Preparation: Michael Brownstein Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2692 row: a column: 18
            High quality sequence stop: 581.
FEATURES   Location/Qualifiers
     source   1..820
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6534234"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_144"
                     /note="Organ: Brain; Vector: pDNR-LIB; Site 1: SfiI
                     (ggccattaatggcc); Site 2: SfiI (ggccgcctcggcc); cDNA made
                     by oligo-dT priming and directionally cloned. 5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3', Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.2-0.5
                     kb size fraction (other fractions present in NIH_MGC 143).
                     Library created in the laboratory of M. Brownstein (NIMH,
                     NIH). Note: this is a NIH_MGC Library."
ORIGIN
        1

Query Match   73.8%; Score 767.4; DB 5; Length 820;
Best Local Similarity   97.6%;
Matches 799; Conservative   0; Mismatches 18; Indels   2; Gaps   2;

Qy   35 CTATTTCCCGGACACAGGCCGTCCACGGTATTGAGTGGTGGTCACAATGGCAGTGAGAAAT 94
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 CTATTTCCCGGACAGGCCGTCCACGGTATTGAGTGGTGGTCACAATGGCAGTGAGAAAT 63

Qy   95 GAATGAAGATTATCCCGTAGAAAATTCACGAGTCTTTAACAGCCCTGGAGAGTCCCTGGG 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 GAATGAAGATTATCCCGTAGAAAATTCACGAGTCTTTAACAGCCCTGGAGAGTCCCTGGG 123

Qy   155 TGCTGTTGGACGACATGCCTAGAGACCATGATGGCTGTTTCTAGAAACGAGTTGTTGCAGA 214
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 TGCTGTTGGACGACATGCCTAGAGACCATGATGGCTGTTTCTAGAAACGAGTTGTTGCAGA 183

Qy   215 GAATGGACCCATTGGAACACAGCAAAGGTGGATTTAGTTCGCATACACCTTAAATTCAAT 274
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 GAATGGACCCATTGGAACACAGCAAAGGTGGATTTAGTTCGCATACACCTTAAATTCAAT 243

Qy   275 GTTTTGGGTTTATTTGGCAACTCAAGGAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGA 334
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 GTTTTGGGTTTATTTGGCAACTCAAGGAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGA 303

Qy   335 ACTGGAAAGAATCAGAGTCTACATGAACACAGAGTTAAAGAAATAACAGACAAGAGAAGGC 394
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 ACTGGAAAGAATCAGAGTCTACATGAACACAGAGTTAAAGAAATAACAGACAAGAGAAGGC 363

Qy   395 TGCCAAGCTGGACAGAGGGCTGCTTCGAGATTTGCTCAAGAAGGCACTCTGGGACCCAA 454
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 TGCCAAGCTGGACAGAGGGCTGCTTCGAGATTTGCTCAAGAAGGCACTCTGGGACCCAA 454

ORIGIN

Query Match   73.9%; Score 768.2; DB 5; Length 871;
Best Local Similarity   95.5%;
Matches 800; Conservative   0; Mismatches 36; Indels   2; Gaps   1;

Qy   57 CACGGTATTGAGTGGTGGTCACAATGGCAGTGAGAAAATGAATGAAGATTATCCCGTAGAA 116
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CACGGTATTGAGTGGTGGTCACAATGGCAGTGAGAAAATGAATGAAGATTATCCCGTAGAA 60

Qy   117 ATTCACGAGTCTTTAACAGCCCTGGAGAGTCCCTGGGTGCTGTTGGACGACATGCTGAAG 176
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATTCACGAGTCTTTAACAGCCCTGGAGAGTCCCTGGGTGCTGTTGGACGACATGCTGAAG 120

Qy   177 ACCATGATGGCTGTTTCTAGAAACGAGTTGTTGCAGAGAGTTGGACCCATTGGAACAAGCA 236
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACCATGATGGCTGTTTCTAGAAACGAGTTGTTGCAGAGAGTTGGACCCATTGGAACAAGCA 180

Qy   237 AAGGTGGGATTTAGTTCTCGCATACACACCTTAAATTCAATGTTTTGGGTTTATTTGGCAACT 296
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAGGTGGGATTTAGTTTCTCGCATACACCTTAAATTCAATGTTTTGGGTTTATTTGGCAACT 240

Qy   297 CAAGGAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGAACTGGAAAGAATCAGAGTCTAC 356
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAAGGAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGAACTGGAAAGAATCAGAGTCTAC 300

Qy   357 ATGAACACAGAGTTAAAGAAATAACAGACAAGAGAAGGCTGCCAGCTGGACAGAGGGCT 416
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATGAACACAGAGTTAAAGAAATAACAGACAAGAGAAGGCTGCCAGCTGGACAGAGGGCT 360

Qy   417 GCTTCGAGATTTGTCAAGAAGCGCTCTGGGAACCCAAAGCACACCAAAAAGTG 476
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GCTTCGAGATTTGTCAAGAAGCGCTCTGGGAACCCAAAGCACACCAAAAAGTG 420

Qy   477 GCTAATAAAAGGGAAAAGCAAACACTAAATCTAATCCTTTCTTTGGTTTTTGATGTACATGTTTTCAAAAAG 536
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GCTAATAAAAGGGAAAAGCAAACACTAAATCTAATCCTTTCTTTGGTTTTTGATGTACATGTTTTCAAAAAG 480

Qy   537 TACATCCTTTTTAATCAGTTAGTTATGTGACCATGTGGTGTTAAAATGGATT 596
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TACATCCTTTTTAATCAGTTAGTTATGTGACCATGTGGTGTTAAAATGGATT 540

Qy   597 CCTTTTGGAATTCATGTATAAAATTAACATTACATTGTGAATACTGAATCTTTTTTTTTTG 656
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCTTTTGGAATTCATGTATAAAATTAACATTACATTGTGAATACTGAATCTTTTTTTTTTG 600

Qy   657 CTGAGAAAGATTAAGTGTGTCTTTGTTGAATTTCATATAAAGCATGATGTGTGTTTAATA 716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CTGAGAAAGATTAAGTGTGTCTTTGTTGAATTTCATATAAAGCATGATGTGTGTTTAATA 660

Qy   717 TTCTAAGATATTCTATAAGCAGTTGTGAAATCCAAATGTTCTCTGTAAACATTGTAGTG 776
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TTCTAAGATATTCTATAAGCAGTTGTGAAATCCAAATGTTCTCTGTAAACATTGTAGTG 720

Qy   777 TTTGAAATGAACAAATGATTATTATGAAGTGTGCTATCTGTAGACTGGAGGTGTAAGGACA 836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GTTGAAATGAACANATGATTATTATGAAGTGTGCTATCTGTAGACTGGAGGTGTAAGGACAT 780

Qy   837 TTTGTTTTCAGTAATGAATGATGAGAGAAA--TACAGTGACTTAAATACCCACTCTGTTTCTGT 892
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||||
Db   781 TTTGTTTTCAGTAATGAATGATGAGAAAATACCCACTCTGTTTCT 838

RESULT 8

```
Qy   364 TGCCAAGCTGGACAGAGGTCGTTCGGAGATTTGTCAAGAACGCACTCTGGGAACCCA 423
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 TGCCAAGCTGGACAGAGGTCGTTCGGAGATTTGTCAAGAACGCACTCTGGGAACCCA 423

Qy   455 ACGAAAAAGCACCAAAAGTGGCTAATAAAGGGAAAAGCAAACACTAATCTTTTGGTTT 514
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 AGCAAAAAGCACCAAAAGTGGCTAATAAAGGGAAAAGCAAACACTAATCTTTTGGTTT 483

Qy   515 TGATGTACAACGTTTTCAAAAAGTACATCCTTTTAATCAGTTGACAAATGTAGTTAATGTGA 574
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 TGATGTACAACGTTTTCAAAAAGTACATCCTTTTAATCAGTTGACAAATGTAGTTAATGTGA 543

Qy   575 CCAATGTGGTGTGTTTAAATGGATTCCTTTGGAATTCATGTATAAAATTACACATTACATTT 634
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 CCAATGTGGTGTGTTTAAATGGATTCATGTATAAAATTACACATTACATTT 603

Qy   635 GTGATACTGAATCTTTTTTTGCTGACGAAAAGATTAAGTTGTTCTTTTGTTTGATTTCATATA 694
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 GTGATACTGAATCTTTTTTTGCTGAGAAAGA-TAAAATTGTTCTTTGTTTGATTTCATATA 662

Qy   695 AAGCATCATGAGTGTTGTTTAAATATTGTAAGATATATTCTATAAGCAGTTGAAATCCAAATG 754
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   663 AAGCATCATGAGTGTTGTTTAAATATTGTAAGATATATTCTATAAGCAGTTGAAATCCAAATG 722

Qy   755 TTCTCTGTAAACATATTTGTAGTGTTTGAAAATGAACAATGATAATATTTAATGAAGTGTGCTATCTG 814
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   723 TTCTCTGTAAACATTTGTAATGTTTGAAAATGGACNATGATTATTATTATGGAAATGTCCTATCTG 782

Qy   815 TAGACCTCGAGGTGTAAGGACATTTGTTGTTTCAGAATGA 853
         ||||||||||||||||||||||||||||||||   |||
Db   783 TAGA-CTGGAGGTGTAAGGACATTTNGTTTCAGAATGA 820
```

LOCUS   CB317510   733 bp   mRNA   linear   EST   04-MAR-2003
DEFINITION   AGENCOURT 12397793 NIH_MGC_165 Mus musculus cDNA clone
    IMAGE:3028471 5', mRNA sequence.
ACCESSION   CB317510
VERSION   CB317510.1   GI:28841745
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 733)
AUTHORS   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE   NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL   Unpublished (1999)
COMMENT   National Institutes of Health, Mammalian Gene Collection (MGC)
    Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-r@mail.nih.gov
    Tissue Procurement: Dr. Leslie L. Heckert
    cDNA Library Preparation: CLONTECH Laboratories, Inc.
    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA Sequencing by: Agencourt Bioscience Corporation
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
    Plate: NDCM49 row: d column: 02
    High quality sequence stop: 584.

FEATURES   Location/Qualifiers
    source   1..733
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /clone="IMAGE:3028471"
    /tissue_type="primary cultures of Sertoli cells"
    /lab_host="DH10B (Tl-phage-resistant)"
    /clone_lib="NIH_MGC_165"
    /note="Organ: testis; Vector: pDNR-LIB; Site 1: SfiI
    (ggccattatggcc); Site 2: SfiI (ggccgcctccggcc); 5' and 3'
    adaptors were used in cloning as follows:5' adaptor
    sequence: 5'-CACGGCGCCATTATGGCC-3', and 3' adaptor sequence:
    5'-ATTCIAGAGGCGGCGGAGGCGGGCGGCCGACATG-dT(30) BN-3' (where B = A,
    C, or G and N = A, C, G, or T). Average insert size 1.4 kb
    (range 0.6-3.5 kb). PCR. This library was enriched for full-length clones and
    was constructed by Clontech Laboratories (Palo Alto, CA).
    PCR. This library was enriched for full-length inserts by
    15/15 colonies contained inserts and 15/15 colonies contained inserts by
    Note: this is a NIH_MGC Library."

ORIGIN

RESULT 9

```
Query Match    697; Conservative   0; Mismatches   4; Indels   0; Gaps   0;
Best Local Similarity   99.4%;
Matches  697; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy    28 GTGCGACCTATTTCCCGGAGGACAGGGGTCCACGGTATTGAGTTGGTCACATGCCAGGTG 87
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2 GGGCGACCTATTTCCCGGAGGACAGGGGTCCACGGTATTGAGTTGGTCACATGCCAGGTG 61

Qy    88 AAGAAATGAATGAAGATTATCCCGTAGAAAATTCACGAGGTCTTTAAACACGCCTGGGAGCT 147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62 AAGAAATGAATGAAGATTATCCCGTAGAAAATTCACGAGGTCTTTAAACAGCCTGGGAGCT 121

Qy   148 CCCTGGGGTGCTGTGACGACAGCAATGCCAATGATTGAGTTAGTTCTGAAACGAGTTGT 207
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 CCCTGGGTGCTGTGACGACAGCAATGCTGAACGACAAAGGGATTAGTTCTGAAACGAGTTGT 181

Qy   208 TGCAGAAGTTGGACCCCATTGGAACAAAGCAAAAGGTGGATTAGTTTTCCATACACCTTAA 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 TGCAGAAGTTGGACCCCATTGGAACAAGCAAAAGGTGGATTAGTTTTCCATACACCTTAA 241

Qy   268 ATTCAAATGTTTTGGGTTTTTATTTGGCCAACTCAAGGAGTTAATCCCAAAGAGCAATCCAAGTGA 327
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 ATTCAAATGTTTTGGGTTTTTATTTGGCCAACTCAAGGAGTTAATCCCAAAGAGCAATCCAAGTGA 301

Qy   328 AGCAGGAACTGGAAAAGAAATCAGAGTTCTACATGAACAGAGTTAAAGAAAATAACAGAGACAAGA 387
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 AGCAGGAACTGGAAAAGAAATCAGAGTTCTACATGAACAGAGTTAAAGAAAATAACAGAGACAAGA 361

Qy   388 AGAAGGCTTGCCAAGCTGGACAGAGGTGCTGCTTCGAGATTGTCAAGAAGAGGCACTCTGG 447
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 AGAAGGCTTGCCAAGCTGGACAGAGGTGCTGCTTCGAGATTGTCAAGAAGAGGCACTCTGG 421

Qy   448 AACCCAAACAAGAAAAAGCAACCAAAAGTGGCTAATAAAGGGAAAAGCAAACACTAATCTT 507
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 AACCCAAAGCAAAAAGCACCAAAAGTGGCTAATAAAGGGAAAAGCAAACACTAATCTT 481

Qy   508 TTGGTTTTGATGTACATGTTTTCAAAAAGTACACCCTTTTAATCAGTTTACAAATGTAGT 567
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482 TTGGTTTTGATGTACATGTTTTCAAAAAGTACATCCTTTTAATCAGTTTACAAATGTAGT 541

Qy   568 TATGTGACCAATGTGGTGTGTTTAAATGGATTCCTTTGGAATTCATGTATGTAAAATTACACAT 627
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 TATGTGACCAATGTGGTGTGTTTAAATGGATTCATGTATAAAATTACACAT 601

Qy   628 TACATTTGTGATACTGAATCTTTTTTTTTGCTGAGAAAGAGATTAAGTTGTTCTTTGTTTGATTT 687
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 TACATTTGTGATACTGAATCTTTTTTTTTGCTGAGAAAGAGATTAAGTTGTTCTTTGTTTGATTT 661

Qy   688 TCATATAAAAGCATCATGAGTGTTGTTTAAATATTGTAAGATATT 728
         |||||||||||||||||||||||||||||||||||||||||||||
Db   662 TCATATAAAAGCATCATGATGTGTTGTTTAAATATTGTAAGATATT 702
```

LOCUS   CB574390   731 bp   mRNA   linear   EST   02-APR-2003
DEFINITION   AGENCOURT 12795777 NIH_MGC_165 Mus musculus cDNA clone
    IMAGE:3027904 5', mRNA sequence.
ACCESSION   CB574390
VERSION   CB574390.1   GI:29493920
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 731)

AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Dr. Leslie L. Heckert
          cDNA Library Preparation: CLONTECH Laboratories, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: NDCM34 row: g column: 23
          High quality sequence stop: 608.

FEATURES   Location/Qualifiers
     source   1..731
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30279046"
                     /tissue_type="primary cultures of Sertoli cells"
                     /clone_lib="NIH_MGC_165"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /note="Organ: testis; Vector: pDNR-LIB; Site 1: SfiI
                     (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); 5' and 3'
                     adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
                     (range 0.6-3.5 kb). 15/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA).
                     Note: this is a NIH_MGC Library."

ORIGIN

Query Match   66.5%; Score 692; DB 6; Length 731;
Best Local Similarity 98.6%; Pred. No. 1.6e-145;
Matches 698; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy   1 CAGAAGCCGTGTCATGGCGTCATCATCGTGCGGACCTATTTCCCGGAGACCAGGCGTCCACG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24 CAGAAGCCGTGTCATCATGGCGTCATCATCGTCGGACCTATTTCCCGAGACCAGGCGTCCACG 83

Qy   61 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAAGATTATCCCGTAGAAATTC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   84 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAAGATTATCCCGTAGAAATTC 143

Qy   121 ACCGAGTCTTTAACAGCCCTGGAGAGCTCCCTGGGTGCTGTGGACCACATGCTGAAGACCA 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   144 ACCAGTCTTTAACAGCCCTGGAGAGCTCCCTGGGTGCTGTGGACCACATGCTGAAGACCA 203

Qy   181 TGATGGCCTGTTCTTCAGAAACGAGTTGTTGGACCCCATTGGAACCAAGCAAAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   204 TGATGGCCTGTTCTTCAGAAACGAGTTGTTGGACCCCATTGGAACCAAGCAAAGG 263

Qy   241 TGGATTTAGTTCTGCATACACACCTTAAATTCAAATTTGGGTTTATTGGGCAACTCAAG 300
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   264 TGGATTTAGTTCTGCATACACACCTTAAATTCAAGTTTGGGTTTATTGGGCAACTCAAG 323

Qy   301 GAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGAACTGGAAAGAATCAGAGTCTACATGA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   324 GAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGAACTGGAAAGAATCAGAGTCTACATGA 383

Qy   361 ACAGAGTAAAGAAAATAACAGACAGAAGAGGAAGAGGCTGCCAAGCTGGACGAGGTGCTT 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   384 ACAGAGTAAAGAAAATAACAGACAGAAGAGGAAGAGGCTGCCAAGCTGGACGAGGTGCTT 443

Qy   421 CGAGATTTGTCAAGAAGGCACTCTGGGAACCAAAAAGCACACCAAAAGTGGGCTA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   444 CGAGATTTGTCAAGAAGGCACTCTGGGAACCAAAAAGCACACCAAAAGTGGGCTA 503

Qy   481 ATAAAAGGGAAAAGCAAACAACTAATCTTTTGGGTTTTGATGTACANGTTTTCAAAAAGTACA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   504 ATAAAAGGGAAAAGCAAACAACTAATCTTTTGGGTTTTGATGTACANGTTTTCAAAAAGTACA 563

Db   564 TCCTTTTTAATCAGTTTACAATGTAGTTATGTGACCAATGTGGTGTTTGTTTAAAATGGATTCCTT 623

Db   624 TTGGAATTCATCTATAAAATTTACACATTACATTTGTCATACTGGATCTTTTTTTTTGCTCAG 683

Qy   661 GAAAGATTAAGTTGTCTTTGTTGATTTTCATATAAAGCCATCATGATGT 708
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   684 AAAAGATTAAGTTGTCTTTGTTGATTTTCATATAAAGCCATCATGATGT 731

RESULT 10
CB598466
LOCUS   CB598466   715 bp   mRNA   linear   EST 03-APR-2003
DEFINITION AGENCOURT 12765813 NIH_MGC_178 Mus musculus cDNA clone
            IMAGE:30297772 5', mRNA sequence.
ACCESSION   CB598466
VERSION   CB598466.1 GI:29516322
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 715)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
            cDNA Library Preparation: Michael Brownstein Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM59 row: d column: 05
            High quality sequence stop: 590.

FEATURES   Location/Qualifiers
     source   1..715
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30297772"
                     /note="Organ: lung and heart; Vector: pDNR-LIB; Site 1:
                     SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA
                     made by oligo-dT priming and directionally cloned. 5' and
                     3' adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3', Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.5 kb
                     size fraction. Library created in the laboratory of M.
                     Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match   66.0%; Score 686; DB 6; Length 715;
Best Local Similarity 99.3%; Pred. No. 3.6e-144;
Matches 689; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy   1 CAGAAGCCGTGTCATGGCGTCATCATCGTGCGGACCTATTTCCCGGAGACCAGGCGTCCACG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22 CAGAAGCCGTGTCATGGCGTCATCATCGTCGGACCTATTTCCCGAGACCAGGCGTCCACG 81

Qy   61 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAAGATTATCCCGTAGAAATTC 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAAGATTATCCCGTAGAAATTC 141

```
RESULT 11
AI875855/c
LOCUS       AI875855        781 bp      mRNA    linear    EST 21-JUL-1999
DEFINITION  uj67b10.x1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1924987 3', similar to TR:O35473 O35473 SMALL UNIQUE NUCLEAR
            RECEPTOR CO-REPRESSOR. [1] ;, mRNA sequence.
ACCESSION   AI875855
VERSION     AI875855.1  GI:5549904
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other ESTs: uj67b10.y1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:981279
```

```
Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 517.
```

```
ORIGIN
Query Match           65.3%;  Score 678.8;  DB 1;  Length 781;
Best Local Similarity 95.9%;  Pred. No. 1.5e-142;
Matches 750; Conservative 0; Mismatches 27; Indels 5; Gaps 5;
```

```
Qy  121 ACGAGTCATTTAAACAGCCTTGGAGAGCTCCCTGGGTGGTGGTGGTGTGGACGACATGCTGAGACCA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142 ACGAGTCATTTAAACAGCCCTGGAGAGCTCCCTGGGTGGTGGTGGTGGACGACATGCTGAGACCA 201

Qy  181 TGATGGCTCGTTTCTTAGAAACGAGTTGTTGTTCGAGAAGTTGGACAACAAGCAAAGG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202 TGATGGCTCGTTTCTTAGAAACGAGTTGTTGTTCGAGAAGTTGGACAACAAGCAAAGG 261

Qy  241 TGGATTTTAGTTTTCTGCATACACCTTAAATTCAATGTTTTGGGCAACTCAAG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262 TGGATTTTAGTTTTCTGCATACACCTTAAATTCAATGTTTTGGGCAACTCAAG 321

Qy  301 GAGTTAATCCCAAAAGACGATCCAGTGAAGCAGGAGAGAAGAATCAGAGAGTCTTACCATGA 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322 GAGTTAATCCCAAAAGACGATCCAGTGAAGCAGGAGAACTGGAAAGAATCAGAGAGTCTTACCATGA 381

Qy  361 ACAGAGTTAAACAAAAATAACACACAAGAAGAGGGCTGCCAAGCTGGACGAGGTGCTGCTT 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  382 ACAGAGTTAAAGAAATAACACAGCAACAAGAAGAGGGCTGCCAAGCTGGACGAGGTGCTGCTT 441

Qy  421 CGAGATTTGTCAAGAGGCCACTCCTGGGAACCCAAAGCGAACACCCAAAAAGTGGGCTA 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  442 CGAGATTTGTCAGGAACGCACTCTCGGGAACCCAAAGCAAAAAGCAACACCCAAAAAGTGGGCTA 501

Qy  481 ATAAAAGGGAAAAAGCAAACAACCAAATCTTTTGGTTTTTGATCGTACATGTTTCAAAAAGTACA 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  502 ATAAAAGGGAAAAAGCAAACAACTAAATCTTTGGTTTTTTGATCGTACATGTTTCAAAAAGTACA 561

Qy  541 TCCTTTTTAATCAGTTTACAAATGTAGTAGTGACCATCTGTTGTGTTTTAAATCGGATTCCTT 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  562 TCCTTTTTAATCAGTTTCACAATGTAGTTATGTGACCATGTGTTGTTTAAATGGATTCCTT 621

Qy  601 TTGGAATTCATGTATAAAATTTACACATTACATTTGTGATACTCGAATCTTTTTTTTGCTGA 660
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  622 TTGGAATTCATGTATAAAATTTACACATTACATTTGTGATACTCGAATCTTTTTTTTTGCTGA 681

Qy  661 GAAAGATTAAGTTGTCTTTGTTGATTTTCATTATA 694
        |||||||||||||||||||||||||||||||||||
Db  682 GAAAGATTAAGTTGTCTTTGTTGATTTTCAAATA 715
```

```
Qy  249 GTTTCTGCATACACCTTAAATTCAATGTTTTGGGTTTTATTTGGCAACTCAAGGAGTTAAT 308
        ||||||||||||   ||||||||||||||||||||| ||||||||||||||||||||||
Db  781 GTTTCGGGCAAACACCTTAAATTCAATGTTTTGGGTTTATTTGGCACTCAAGGAGTTTAAT 722

Qy  309 CCCAAAAGACGCATCCAGTGAAGCAGGAGAGAAGAACTGGAAAGATCTGAAAAGTGACACAGAGTT 368
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 CCCAAAAGAGCATCCAGTGAAGCAGGAGAACTGGAAAGAATCAGAGTCTTACATGAACCAGAGTT 662

Qy  369 AAAGAAAATAACACAGCAAGAAGAGGGCTGCCAAGCTGGACAGAGGTGCTGCTTCGGATTT 428
        ||||||||||||||||||||||||||||||||||||||||  |||||||||||||||||||
Db  661 AAAGAAAATAACAGCAAGAAGAGGGCTGCCAAGCTGGACGAGGTGCTGCTTCGAGATTT 602

Qy  429 GTCAAGAA-GGCACTCTGGGAACCCAAAGCAAAAAGCACA-CCAAAAGTGGCTAATAAAG 486
        |||||||| |||||||||||||||||||||||||||||   ||||||||| |||||||||
Db  601 GTCAAGAACGCCACTCTCTGGGAACCCAAAGCAAAAAGCACACCCCAAAAAGTGGGCTAATAAAG 542

Qy  487 GGAAAAAGCAAACACTAAATCATTTCTTTTGGTTTTTGATGTACATGTTTCAAAAAGTACATCCTTTT 546
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GGAAAAAGCAAACACTAAATCTTTGGTTTTTGATGTACATGTTTCAAAAAGTACATCCTTTT 482

Qy  547 TTAATCAGTTTACAATGTAGTTATGTGACCATGTGGTGTTTAAATGGATTCCTTTTTCTTTTGAA 606
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 TTAATCAGTTTACAATGTAGTTATGTGACCATGTGGTGTTTAAATGGATTCCTTTTTGGAA 422

Qy  607 TTCATGTATAAAATTACACATTACATTGTGATACTGAATCTTTTTTTTGCTGAGAAAGA 666
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 TTCATGTATAAAATTTACACATTACATTTGTGATACTCGAATCTTTTTTTTGCTGAGAAAGA 362

Qy  667 TTAAGTTGTCTTTGTTGATTTTCATAAAGCATCATCATGATGTGTTTTAATGTTGTAAGATA 726
        |||||||||||||||||||||||| |||| |||||||||||| |||||||||||||||||||
Db  361 TTAAGTTGTCTTTGATGTTTCATATATAAAGCATCATGATGTGTTTTAATATTGTAAGATA 302

Qy  727 TTCTATAAGCAGTTGTGAAAATCCAAATGTTCTCTGTAAAACATTTGTAGTTGTGTTGAAATGA 786
        |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  301 TTCTATAAGCAGTTGTGAAATCCAAATGTTCTCTGTAAAACATTTGTAGTTGTGTTGAAATGA 242

Qy  787 ACAATGATATTATGAAGTGTGCTATCTGTAGACCTCGAGGTGTAAGGACATTGTTTCA 846
        ||||||||||||||||||||||||||||||||||||||||||||||   |||||||
Db  241 ACAATGATATTATGAAGTGTGCTATCTGTAGA-CTGGAGGTGTAAGGACATTGTTTCA 183
```

```
RESULT 12

LOCUS       BB612968                685 bp    mRNA    linear    EST 31-AUG-2001
DEFINITION  BB612968 RIKEN full-length enriched, 10 day neonate skin Mus
            musculus cDNA clone 4732419J08 5', mRNA sequence.
ACCESSION   BB612968
VERSION     BB612968.1  GI:15395380
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 685)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,T., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
            RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
            Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Yamanaka,H., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K.,
            Arakawa,T., Ishii,Y. and Hayashizaki,Y.
            Mapping of 19022 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, L72-L86 (2001
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
            e mouse tissues.
```

```
                                                      source          1..685
                                                                      /organism="Mus musculus"
                                                                      /mol_type="mRNA"
                                                                      /strain="C57BL/6J"
                                                                      /db_xref="taxon:10090"
                                                                      /clone="4732419J08"
                                                                      /sex="mixed"
                                                                      /tissue_type="skin"
                                                                      /dev_stage="10 days neonate"
                                                                      /clone_lib="RIKEN full-length enriched, 10 day neonate
                                                                      skin"
                                                                      /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                                                                      prepared and sequenced in Mouse Genome Encyclopedia
                                                                      Project of Genome Exploration Research Group in Riken
                                                                      Genomic Sciences Center and Genome Science Laboratory in
                                                                      RIKEN. Division of Experimental Animal Research in Riken
                                                                      contributed to prepare mouse tissues. 1st strand cDNA was
                                                                      primed with a primer [5'
                                                                      GAGAGAGAGAGGANTCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3']. cDNA was
                                                                      prepared by using trehalose thermo-activated reverse
                                                                      transcriptase and subsequently enriched for full-length by
                                                                      cap-trapper. cDNA went through one round of normalization
                                                                      to Rot = 10.0 and subtraction to Rot = 100.0. Second
                                                                      strand cDNA was prepared with the primer adapter of
                                                                      sequence [5' GAGAGAGAGATTCGGCACGAGG 3']. cDNA was
                                                                      3']. cDNA was cloned into the XhoI and BamHI sites.
                                                                      Vector: a modified pBluescript KS(+) after bulk excision
                                                                      from Lambda FLC I"

ORIGIN

   Qy    Query Match
   Db    Best Local Similarity   65.2%;
         Matches 680;  Conservative   99.6%;   Score 678.2;   DB 2;   Length 685;
                                    0;  Mismatches   3;   Indels   0;   Gaps   0;
```

Alignment section:

```
Qy   847 GTAATGATGAGAAAATACACAGTGACTTAAATACCCACACTCTG-TTTCTGTTCAGTTAGTTCAA 905
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 GTAATGATGAGAAAATACAGTGACTTAAATACCCACTCTGTTTTCTGTTCAGTTAGTTCAA 123

Qy   906 CATGTTTCGTGATTTTTTTTTTTTTTTTTTTGAGTAGTTA 964
         |||||||||
Db   122 CATGTTTCGTGATTTTTTTTTTTTTTTTTTTGAGTAATTGCTGTGTTCTGATATTCAAAGT-CAAAA 63

Qy   965 TTGAAACCTTAAGGCTGTACTTTAATTCTTCATGTTCCATTAAAATAAAATAAAATGTTCTCAT 1024
         ||||||||||||||||||||||||||||||||||||
Db    62 TTGAAACCTTAAGGCTGTACTTTAGTTCTTCATGTTCCATTAAAATAAAATGTTCTCAT 3

Qy  1025 TA 1026
         ||
Db     2 TA 1
```

ORIGIN alignment:

```
Qy    22 ATCATCGTGCGACCTATTTCCCGGAGACAGGCGTCCACGGTAGTTGAGTTGGTCACAATGG 81
Db     2 ATCATCGTGCGACCTATTTCCCGGAGACAGGCGTCCACGGTATTGAGTTGGTCACAATGG 61

Qy    82 CAGGTGAGAAATGAATGAAGATTATCCCGTAGAAAATTCACGGAGTCTTTAACAGCCCTGG 141
Db    62 CAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAAATTCACGGAGTCTTTAACAGCCCTGG 121

Qy   142 AGAGCTCCCCTGGGGTGCTGTGGACGACATGCTGAAGACCATGGCTGTTTCTTAGAAACG 201
Db   122 AGAGCTCCCCTGGGGTGCTGTGGACGACATGCTGAAGACCATGGCTGTTTCTTAGAAACG 181

Qy   202 AGTTGTTGCAGAAGTTGGACCCCATTGGAACAAGCAAAAGGTGGATTTAGTTGTCGCATACA 261
Db   182 AGTTGTTGCAGAAGTTGGACCCCATTGGAACAAGCAAAAGGTGGATTTAGTTGTCGCATACA 241

Qy   262 CCTTAAAATTCAAATGTTTGGGTTTATTGGCAACTCAAGGAGTTAATCCCAAAGAGCCATC 321
Db   242 CCTTAAAATTCAAATGTTTGGGTTTATTGGCAACTCAAGGAGTTAATCCCAAAGAGCATC 301

Qy   322 CAGTGAAGCACGAAGTGGAAAGAATCAGAGTCTACACAGAGTTAAAGAAATAACAG 381
Db   302 CAGTGAAGCACGAAGTGGAAAGAATCAGAGTCTACACAGAGTTAAAGAAATAACAG 361

Qy   382 ACAAGAAGAAGGACTGCCAAGCTGGACAGAGCTGGTGCTTGCTTCGAGATTTGTCAAGAAGCGAC 441
Db   362 ACAAGAAGAAGGACTGCCAAGCTGGACAGAGCTGGTGCTTGCTTCGAGATTTGTCAAGAAGCGAC 421

Qy   442 TCTGGGAAACCCAAACGAAAAAAGCACACCAAAAGGTGGCTAATAAAGGGAAAAAGCAAACACT 501
Db   422 TCTGGGAAACCCAAACGAAAAAAGCACACCAAAAGGTGGCTAATAAAGGGAAAAAGCAAACACT 481

Qy   502 AATCTTTTGGTTTGATTGACATGTTTTCAAAAAGTACATCCTTTTTAATCAGTTTACA 561
Db   482 AATCTTTTGGTTTGATTGACATGTTTTCAAAAAGTACATCCTTTTTAATCAGTTTACA 541

Qy   562 TGTAGTTATGTGACCATGTGGTGTTAAATGGATTCCTTTTTGGAATTCATGTATAAATTT 621
         ||||||||
```

RESULT 13

```
LOCUS       BQ746434     690 bp    mRNA    linear   EST 17-JUL-2002
DEFINITION  UI-M-ER0-bxp-f-16-0-UI.r1 NIH_BMAP_ER0 Mus musculus cDNA clone
            IMAGE:5711247 5', mRNA sequence.
ACCESSION   BQ746434
VERSION     BQ746434.1  GI:21893221
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
FEATURES             Location/Qualifiers
     source          1..690
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5711247"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 15.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_ER0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library
                     sequence located between the Not I site and the polyA
                     tail, is GTGCGTGGAA. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Institutes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
                     Seq primer: pYX-5.
```

ORIGIN

```
Query Match     63.6%; Score 661.6; DB 5; Length 690;
Best Local Similarity  99.1%; Pred No.1.1e-138;
Matches  686; Conservative  0; Mismatches  4; Indels  2; Gaps  2;

Qy    57 CACGGTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAATGAAGAAGATTATCCCGTAGAA 116
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 TGTAGTTATGTGATACTGAATCTTTTTTTTGCTGAGAAAGATTAAGTTGTCTTTTGT 601

Qy   622 ACACATTACATTTGTGATACTGAATCTTTTTTTTGCTGAGAAAGATTAAGTTGTCTTTGT 681
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 ACACATTACATTTGTGATACTGAATCTTTTTTTGCTGAGAAAGATTAAGTTGTCTTTGT 661

Qy   682 TGATTTTTCATTATAAAGCATCATG 704
         |||||||||||||||||||||||||
Db   662 TGATTTTTCATTATAAAGCATCATG 684
```

RESULT 14

```
LOCUS       CD352734     671 bp    mRNA    linear   EST 15-JUL-2003
DEFINITION  UI-M-GL0-cfy-h-18-0-UI.r1 NIH_BMAP_GL0 Mus musculus cDNA clone
            IMAGE:30359321 5', mRNA sequence.
ACCESSION   CD352734
VERSION     CD352734.1  GI:31145235
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be found
            at
            http://genome.uiowa.edu/distribution/mousefl.html
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
```

```
Db     1 CACGGTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAATGAAGAAGATTATCCCGTAGAA 60

Qy   117 ATTCACGGAGTCTTTAACAGGCCCTGGGAGCAGAGCTCCCTGGGTGGTGGGAGCGGACGATGCTGAAG 176
         |||||||||||||||  ||||| ||||| ||||||||||||||||||||||||  ||||  ||||||
Db    61 ATTCACGGAGTCTTT-ACAGCCCTGGAGAGCTCCCTGGGTGGTGGGACGACATGCTGAAG 119

Qy   177 ACCATGATGGCTGCTGTTCTTAGAAACGAGTTGTTGCAGAAGTTGGACCCATTGGAACAAGCA 236
         ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   120 ACCATGATGGCTGCTTGTTCTTAGAAACGAGTTGTTGCAGAAGTTGGACCCATTGGAACAAGCA 179

Qy   237 AAGGTGGATTTAGTTTCTGCAATACACCTTAAATTCAATATTGTTTGGGTTTATTGGGCAACT 296
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   180 AAGGTGGATTTAGTTTCTGCAATACACCTTAAATTCAATATTGTTTGGGTTTATTGGGCAACT 239

Qy   297 CAAGGAGTTAATCCCAAAAGACGCATCCAGTGAAGCAGGAAAGCAGAAGTCGAGAGTCTAC 356
         |||||||||||||||||||||| ||||||||||||||||||||  |||||||||||||||||
Db   240 CAAGGAGTTAATCCCAAAAGAGCATCCAGTGAAGCAGGAAACTGGAAGAAATCAGAGTCTAC 299

Qy   357 ATGAACAGAGTTAAAGAAAAATAACAGACAAGAAGAAGAAGGCTGCCAAGCTGGACAGAGGTGCT 416
         |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   300 ATGAACAGAGTTAAAGAAAATAACAGACAAGAAGAAGAAGGCTGCCAAGCTGGACAGAGGTGCT 359

Qy   417 GCTTCGAGATTTGTCAAGAAGGCACTCTGGGAACCCAAAACGAAAAAGCACCACCAAAAGTG 476
         |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   360 GCTTCGAGATTTGTCAAGAAGGCCACTCTGGGAACCCAAAACGAAAAAGCACCACCAAAAGTG 419

Qy   477 GCTAATAAAGGGAAAACAAACAACACTAATCTTTCTTTGTTTTGTTGTTGTTTTTCAAAAG 536
         |||||||||||||||| |||||||||||||  ||||   ||||| |||| || |||||||||
Db   420 GCTAATAAAGGGAAAAGCAAACAACACTAATC-TTTGGTTTGTTGATGTACATGTTTTTCAAAAG 478

Qy   537 TACATCCTTTTTTTAATCAGTTTACAATGTAGTTGTATGTGACCATGTGGTGTTTACATGGATT 596
         |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   479 TACATCCTTTTTTAATCAGTTTACAATGTAGTTGTATGTGACCATGTGGTGTTTACATGGATT 538

Qy   597 CCTTTTGGAATTCATGTATATAAAATTTCACATTACATTTGTGATACTGAATCTTTTTTTTG 656
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539 CCTTTTGGAATTCATGTATATAAAATTTCACATTACATTTGTGATACTGAATCTTTTTTTG 598

Qy   657 CTGAGAAAGATTAAGTTGTCTTTGTTGATTTTTCATATAAAGCATCATGATGTGTTTAATA 716
         |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   599 CTGAGAAAGATTAAGTTGTCTTTGTTGATTTTTCATATAAAGCATCATGATGATGTGTTTAATA 658

Qy   717 TTGTAAGATATTCTATTAAGCAGTTGTGAAATC 748
         ||||||||||||||||||||||||||||||| ||
Db   659 TTGTAAGATATTCTATTAAGCAGTTGTGAAAATC 690
```

FEATURES
source   Seq primer: pYX-5.
            Location/Qualifiers
            1..671
            /organism="Mus musculus"
            /strain="C57BL/6"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:3039321"
            /tissue_type="whole brain"
            /dev_stage="1, 5 and 15 days newborn"
            /lab_host="DH10B (T1 phage resistant)"
            /clone_lib="NIH BMAP G10"
            /note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
            Site_2: Not I; The library was constructed according
            Bonaldo, Lennon and Soares, Genome Research,
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel.First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site.Double strand cDNA was size
            selected according to mRNA size fraction, ligated with
            EcoR I adaptor, digested with NotI and then cloned
            directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is CGAACTGAAT. This library was created for the University
            Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
            program coordinator."

ORIGIN
Query Match   62.9%;   Score 654.2; DB 6;   Length 671;
Best Local Similarity 99.4%;   Pred. No. 5.4e-137;
Matches 667; Conservative   0; Mismatches   3; Indels   1; Gaps   1;

Qy   74 CACAATGGCAGGTGAAGAAATGAATGAAGAATTATCC-CGTAGAAATTCACGAGTCTTTAA 132
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1 CACAATGGCAGGTGAAGAAATGAATGAAGAATTATCCNCGTAGAAATTCACGAGTCTTTAA 60

Qy   133 CAGGCCCTGGAGAGCTCCCTGGGTGCTGTGGACGACATGTGAAGACCATGATGGCTGTTT 192
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAGGCCCTGGAGAGCTCCCTGGGTGCTGTGGACGACATGTGAAGACCATGATGGCTGTTT 120

Qy   193 CTAGAAACCAGTTGTTTGCAGAAGTTGGACCCATTCGAACAAGCAAAGGTGGATTTAGTTT 252
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTAGAAACCAGTTGTTTGCAGAAGTTGGACCCATTGGAACAAGCAAAGGTGGATTTAGTTT 180

Qy   253 CTGGCATACACCTTAAATTCAAATGTTTGGGTTTATTGGCCAACTCAAGGAGTTAATCCCA 312
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGGCATACACCTTAAATTCAAATGTTTGGGTTTATTGGCAACTCAAGGAGTTAATCCCA 240

Qy   313 AAGAGCATCCAGTGAAGCAGGAACTGGGAAAAGAATCAGAGTCTACATGAACAGAGTTAAG 372
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AAGAGCATCCAGTGAAGCAGGAATCAGAGTCTACATGAACAGAGTTAAAG 300

Qy   373 AAATAACAGACAACAAGAAGGAGCTGCCAAGCTGGACAGAGGTGCTGCTTCGAGATTTGTCA 432
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAATAACAGACAACAAGAAGGAGCTGCCAAGCTGGACAGAGGTGCTGCTTCGAGATTTGTCA 360

Qy   433 AGAAGGCACTCTCGGGAACCCAAACGAAAAAGCACACAAAAGTGGCTAATAAGGGAAA 492
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGAACGCACTCTCGGGAACCCAAACGAAAAAGCACACAAAAGTGGCTAATAAGGGAAA 420

Qy   493 GCAAACACTAATCTTTGGTTTTGATGATGTACATGCATCCTGTTTTTAATC 552
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GCAAACACTAATCTTTGGTTTTGATGATGTACATGCATCCTTTTTAATC 480

Qy   553 AGTTTACAAATGTAGTTAGTTGATGACCATGTGTCGTTGTTTAAATGGAATTCATG 612
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGTTTACAAATGTAGTTATGTGACCATGTGTTTTAAATGGAATTCATG 540

Qy   613 TATAAATTTACACATTACATTGTTGTCATGTGAATCTTTTTTTTTGCTGAGAAAGATTAAGT 672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TATAAATTTACACATTACATTGTTGTCATGTGAATCTTTTTTTTTGCTGAGAAAGATTAAGT 600

RESULT 15
LOCUS   CB320669   772 bp   mRNA   linear   EST 04-MAR-2003
DEFINITION AGENCOURT 12427115 NIH_MGC_165 Mus musculus cDNA clone
            IMAGE:30285684 5', mRNA sequence.
ACCESSION   CB320669
VERSION   CB320669.1 GI:28844904
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 772)
  AUTHORS   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Leslie L. Heckert
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM51 row:1 column: 13
            High quality sequence stop: 578.

FEATURES   Location/Qualifiers
     source   1..772
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30285684"
                     /tissue_type="primary cultures of Sertoli cells"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_165"
                     /note="Organ: testis; Vector: pDNR-LIB; Site_1: SfiI
                     (ggccatatggcc); Site_2: SfiI (ggccgcctcggcc); 5' and 3'
                     adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
                     (range 0.6-3.5 kb). 15/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA).
                     Note: this is a NIH_MGC Library."

ORIGIN
Query Match   62.7%;   Score 651.6; DB 6;   Length 772;
Best Local Similarity 98.4%;   Pred. No. 2.1e-136;
Matches 679; Conservative   0; Mismatches   9; Indels   2; Gaps   2;

Qy   1 CAGAAGCCCGTGTCATGGCGTCATCATCGTCGACCTATTCCCGAGACAGGCGGTCCACG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 CAGAAGCCCGTGTCATGGCGTCATCGTCGGACCTATTCCCGAGACAGGCGGTCCACG 80

Qy   61 GTATTGAGTTGGGTCACAAATGGCAGTGAAGAAATGAATGAAGAATTATCCCGTAGAAATTC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 GTATTGAGTTGGGTCACAAATGGCAGTGAAGAAATGAATGAAGAATTATCCCGTAGAAATTC 140

Qy   121 ACGAGTCTTTAACAGGCCCTGGAGAGCTCCCTGGGTGCTGTGGACGACATGCTGAAGACCA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 ACGAGTCTTTAACAGGCCCTGGAGAGCTCCCTGGGTGCTGTGGACGACATGCTGAAGACCA 200

TGTCTTCTTTTGTTTTTCATTATAAAAAGCATCATGATGTGTGTTTAATATTGTTAAGTAGGATTATTCTAT 732
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TGTCTTTTGTTTTGTTTTTCATTATAAAAAGCATCATGATGTGTTTAATATTGTTAAGTAGGATTATTCTAT 660

Qy   733 AAGCAGTTGTG 743
             |||||||||||
Db   661 AAGCAGTTGTG 671

```
Qy   181 TGATGGCTGTTTCTTAGAAAACGAGTTGTTGCAGAGTTGGACCCATTGGAACAAGCAAAGG 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 TGATGGCTGTTTCTTAGAAATGAGTTGTTGCAGAGTTGGACCCATTGGAACAAGCAAAAGG 260

Qy   241 TGGATTTAGTTTCTGCATACACCTTAAATTCAAATGTTTTTGGCACACTCAAG 300
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   261 TGGATTTAGTTTCTGCATACACCTTAAATTCAAATGTTTTTGGCACACTCAAG 320

Qy   301 GAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGAACTGGAAAAGAATCAGAGTCTACATGA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 GAGTTAATCCCAAAGAGCATCCAGTGAAGCAGGAACTGGAAAAGAATCAGAGTCTACATGA 380

Qy   361 ACAGAGTTAAAGAAATAACAGACAAGAAAGGCTGCCAAGCTCGGACAGAGGTGCTGCTT 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   381 ACAGAGTTAAAGAAATAACAGACAAGAAAGGCTGCCAAGCTCGGACAGAGGTGCTGCTT 440

Qy   421 CGAGATTTGTCAAGAAGGCACTCTCGGGAACCCAAACGAAAAAGGACACCCAAAAGTGGCTA 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   441 CGAGATTTGTCAAGAACGCACTCTCGGGAACCCAAAGCAAAAGCACACCCAAAAGTGGCTA 500

Qy   481 ATAAAGGGAAAAAGCAAAACACTAAATCTTTTGGTTTTGATGTACAIGTTTTCAAAAAGTACA 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   501 ATAAAGGGAAAAAGCAAAACACTAAATCTTTTGGTTTTGATGTACAIGTTTTCAAAAAGTACA 560

Qy   541 TCCTTTTTTAATCAGTTTACAAGTAGTAGTGATGTGACCATGTGGTGTGTTTAAATGGATTCCTT 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   561 TCCTTTTTTAAATCAGTTTACAAGTAGTTAGTGACCATGTGTGTTTAAATGGATTCCTT 620

Qy   601 TT-GGAATTCATGTATAAATTTACACATTACATTTGTGATACTGAATC-TTTTTTTTGCT 658
         || |||||||||||||||||||||||||||||||||||||| ||||||
Db   621 TTCGGGATTTCATGTATAAATTTACACATTACATTTGTGATACTGAATCTTTTTTTTTGCT 680

Qy   659 GAGAAAGATTAAGTTGTCTTTGTTGATTTTT 688
         ||||||||||||||||||||||||| |||
Db   681 GAAAAAGATTAAGTTGTCTTTGGTGAATTT 710
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: January 4, 2005, 03:16:16

OM nucleic - nucleic search, using sw model

Title: US-09-701-618A-3
Perfect score: 1040
Sequence: 1 cagaagccgtgtcatggcgt.......tcattaacttctgatggaaaa 1040

Scoring table: IDENTITY_NUC
Searched: Gapop 10.0 , Gapext 1.0
Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                             1: gb_ba:*
                             2: gb_htg:*
                             3: gb_in:*
                             4: gb_om:*
                             5: gb_ov:*
                             6: gb_pat:*
                             7: gb_ph:*
                             8: gb_pl:*
                             9: gb_pr:*
                             10: gb_ro:*
                             11: gb_sy:*
                             12: gb_un:*
                             13: gb_vi:*
                             14: gb_vi:*

Search time 4742.51 Seconds
(without alignments)
10370.300 Million cell updates/sec

1040

4526729 seqs, 2364849745 residues

9053458

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1040	100.0	1040	6	BD221635	BD221635 Method of
2	1040	100.0	1040	6	AX009363	AX009363 Sequence
3	1040	100.0	1050	6	MMC1DPROT	MMC1DPROT
4	960.4	92.3	1040	6	X95591	X95591 M.musculus
5	944.6	90.8	2000	10	BC005436	BC005436 Mus muscu
6	767.8	73.8	215616	2	AF031426	AF031426 Mus muscu
c 7	767.8	73.8	224919	2	AC109025	AC109025 Rattus no
c 8	642	61.7	232869	10	AC110304	AC110304 Rattus no
c 9	642	61.7	235365	10	AL603925	AL603925 Mouse DNA
c 10	564.8	54.3	789	10	AC044845	AC044845 Mus muscu
11	467.2	44.9	221055	2	AY302220	AY302220 Cricetulu
12	467.2	44.9	209355	2	AC123163	AC123163 Rattus no
13	433.8	41.7	1156	6	AC097131	AC097131 Rattus no
14	433.8	41.7	1156	6	BD221634	BD221634 Method of
15	431	41.4	1172	6	AX009361	AX009361 Sequence
16	431	41.4	1172	6	AX329809	AX329809 Sequence
17	431	41.4	1172	6	AX779798	AX779798 Sequence
18	431	41.4	1172	6	AX779799	AX779799 Sequence
19	416.2	40.0	1186	9	BC009589	BC009589 Homo sapi
20	413.6	39.8	172079	9	AL356095	AL356095 Human DNA
21	409	39.3	186856	9	AL354750	AL354750 Human DNA
22	409	39.3	195426	2	AC068741	AC068741 Homo sapi
23	398.6	38.3	1168	9	AC016284	AC016284 Homo sapi
24	396.6	38.1	1195	9	BC009584	BC009584 Homo sapi
25	395.6	38.0	1199	9	BC005235	BC005235 Homo sapi
26	391.2	37.6	700	11	BV073361	BV073361 S212P6211
c 27	369	35.5	212115	2	AC116701	AC116701 Mus muscu
c 28	369	35.5	238058	2	AC114654	AC114654 Mus muscu
c 29	326.8	31.4	426	6	AX452900	AX452900 Sequence
c 30	326.8	31.4	426	6	AX452901	AX452901 Sequence
c 31	325.6	31.3	420	6	AX467454	AX467454 Sequence
c 32	325.6	31.3	420	6	AX467455	AX467455 Sequence
c 33	314.6	30.3	420	6	AX452903	AX452903 Sequence
c 34	314.6	30.3	420	6	AX452904	AX452904 Sequence
c 35	298.4	28.7	457	6	CQ704080	CQ704080 Sequence
c 36	263.2	25.3	380	6	AX885372	AX885372 Sequence
c 37	263.2	25.3	380	6	BD024982	BD024982 Sequence
c 38	256.2	24.6	154312	9	AP001767	AP001767 Homo sapi
c 39	256.2	24.6	178728	2	AC087613	AC087613 Homo sapi
40	226	21.7	342	11	BV179135	BV179135 sqm10401
41	221.4	21.3	4018	9	BX647771	BX647771 Homo sapi
c 42	189.2	18.2	728	9	CR353144	CR353144 Gallus ga
c 43	177.6	17.1	155888	9	AC079112	AC079112 Homo sapi
c 44	84.4	8.1	125547	2	AC121162	AC121162 Homo sapi
45	84.4	8.1	207100	10	AL773585	AL773585 Mouse DNA

ALIGNMENTS

RESULT 1
BD221635
LOCUS   BD221635   1040 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Method of inducing apoptosis in cell.
ACCESSION   BD221635
VERSION   BD221635.1 GI:31031405
KEYWORDS   JP 2002517192-A/2.
SOURCE   Mus sp.
  ORGANISM Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1040)
  AUTHORS   Rothbarth,K., Stammer,H., Werner,D. and Nails,P.
  TITLE   Method of inducing apoptosis in cell
  JOURNAL   Patent: JP 2002517192-A 2 18-JUN-2002;
            DEUTSCHES KREBSFORSCHUNGSZENTRUM STIFTUNG DES OFFENTLICHEN RECHTS,
            PETER NAILS
COMMENT   OS   Mus sp. (mouse)
            PN   JP 2002517192-A/2
            PD   18-JUN-2002
            PF   02-JUN-1999 JP 2000552267
            PR   03-JUN-1998 DE 198 24 811.3
            PI   KARSTEN ROTHBARTH,HERMANN STAMMER,DIETER WERNER,PETER NAILS PC
            C12N15/09,A61K38/00,A61K48/00,A61P35/00,C12N15/00,A61K37/02 CC
            Method of inducing apoptosis in cell
FH   Key   Location/Qualifiers
FT   CDS   (78)..(500).
FT   mat_peptide   (78)..(500).
FEATURES   Location/Qualifiers
     source   1..1040
                     /organism="Mus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10095"
ORIGIN

Query Match   100.0%; Score 1040; DB 6; Length 1040;
Best Local Similarity   100.0%; Pred. No. 3.2e-197;
Matches 1040; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CAGAAGCCGTGTCATGGCGTCATCATCGTGCGGACCTTATTTCCCGGAGACAGGCGTCCACG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CAGAAGCCGTGTCATGGCGTCATCATCGTGCGGACCTTATTTCCCGGAGACAGGCGTCCACG 60

```
RESULT 2
AX009363
LOCUS       AX009363       1040 bp    DNA    linear    PAT 06-SEP-2000

DEFINITION  Sequence 3 from Patent WO9963071.
ACCESSION   AX009363
VERSION     AX009363.1  GI:9996664
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Rothbarth,K., Werner,D. and Stammer,H.
  TITLE     Method for triggering apoptosis in cells
  JOURNAL   Patent: WO 9963071-A 3 09-DEC-1999;
            DEUTSCHES KREBSFORSCH (DE); ROTHBARTH KARSTEN (DE); WERNER DIETER
            (DE); STAMMER HERMANN (DE)
FEATURES             Location/Qualifiers
     source          1..1040
                     /organism="Mus sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10095"
     CDS             78..503
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC07451.1"
                     /db_xref="GI:9996665"
                     /translation="MAGEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELLERIRYYMNNRVKEITD
                     KKKAAKLDRGAASRFVKKALWEPFKRKSTPKVANKGKSHI"
ORIGIN

Query Match       100.0%;   Score 1040;  DB 6;  Length 1040;
Best Local Similarity 100.0%;
Matches 1040;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
                          Pred. No. 3.2e-197;
```

(alignment of Query (Qy) against Database sequence (Db), 100% identity over 1040 bp, in 60-nucleotide blocks)

RESULT 3

```
LOCUS       MMC1DPROT               1050 bp    mRNA    linear   ROD 20-FEB-1998
DEFINITION  M.musculus mRNA for C1D protein.
ACCESSION   X95591
VERSION     X95591.1  GI:1185124
KEYWORDS    C1D gene; C1D protein; DNA-binding protein.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Nehls,P., Keck,T., Greferath,R., Spiess,E., Glaser,T.,
            Rothbarth,K., Stammer,H. and Werner,D.
  TITLE     cDNA cloning, recombinant expression and characterization of
            polypeptides with exceptional DNA affinity
  JOURNAL   Nucleic Acids Res. 26 (5), 1160-1166 (1998)
  MEDLINE   98136194
   PUBMED   9469821
REFERENCE   2  (bases 1 to 1050)
  AUTHORS   Werner,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-1996) D. Werner, Dt. Krebsforschungszentrum,
            Biochemistry of the Cell, Im Neuenheimer Feld 280-0225, D-69120
            Heidelberg, FRG
FEATURES             Location/Qualifiers
     source          1..1050
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="C1D-580"
                     /cell_line="Ehrlich HD34"
                     /cell_type="tumor cell"
                     /tissue_type="ascites tumor"
                     /clone_lib="lambda gt10 (CBA10S)"
     gene            1..1050
                     /gene="C1D"
     CDS             78..503
                     /gene="C1D"
```

```
                     /note="DNA-binding protein"
                     /codon_start=1
                     /product="C1D protein"
                     /protein_id="CAA64844.1"
                     /db_xref="GI:1185125"
                     /db_xref="GOA:Q61368"
                     /db_xref="TrEMBL:Q61368"
                     /translation="MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERLRVVNMRVKEITD
                     KKKAAKLDRGAASRFVKKALWEPRKRKSTPRVVANGKSKH"
ORIGIN
```

Query Match   100.0%; Score 1040; DB 10; Length 1050;
Best Local Similarity 100.0%; Pred. No. 3.2e-197;
Matches 1040; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
polyA_signal    1009..1014
                /gene="C1D"
```

```
Qy      1 CAGAGACCGTGTCATGTCGGTCATCATCGTGTCGGGACCTATTCCCGGAGACCGTCACG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CAGAGACCGTGTCATGTCGGTCATCATCGTGTCGGGACCTATTCCCGGAGACCGTCACG 60

Qy     61 GTATTGAGTTGGGTCACAAATGGCCAGGTGAAGAAATGAATGAAGATTATCCGTGAAATTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTATTGAGTTGGGTCACAAATGGCCAGGTGAAGAAATGAATGAAGATTATCCGTGAAATTC 120

Qy    121 ACGAGTCTTTAACACGCCCTGAGAGCTCCCGGGACGTGCTGTGGACGACATGCTGAAGACCA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACGAGTCTTTAACACGCCCTGAGAGCTCCCGGGACGTGCTGTGGACGACATGCTGAAGACCA 180

Qy    181 TGATGGCTGTGTTTCTAGAAAGGCAGTTGTTGCAGAAGTTGGGACCCATTGGAACAAGCAAAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGATGGCTGTGTTTCTAGAAAGGCAGTTGTTGCAGAAGTTGGGACCCATTGGAACAAGCAAAGG 240

Qy    241 TGGATTGAGTTGTTCTGCATACACCTTAAATTCAAATGTTTGGGTTTATTGGCAACTCAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGGATTGAGTTGTTCTGCATACACCTTAAATTCAAATGTTTGGGTTTATTGGCAACTCAAG 300

Qy    301 GAGTTAATCCCAAAGAGCCATCCAGTGAAGCAGGAACTGGAAAGAAATCAGAAGTCTCACATGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GAGTTAATCCCAAAGAGCCATCCAGTGAAGCAGGAACTGGAAAGAAATCAGAAGTCTCACATGA 360

Qy    361 ACAGAGTTAAAGAAATAAACAGACAAGAAGGCTGCCAAGCCTGGACAGAGGTGCTGCTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACAGAGTTAAAGAAATAAACAGACAAGAAGGCTGCCAAGCCTGGACAGAGGTGCTGCTT 420

Qy    421 CGAGATTTGTCAAGAAGGCACTCTCTGGGGAACCCAAACGAAAAAGCACACCAAAAGTGGCTA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGAGATTTGTCAAGAAGGCACTCTCTGGGGAACCCAAACGAAAAAGCACACCAAAAGTGGCTA 480

Qy    481 ATAAAACGGAAAAAGCAAACACTAATCTTTTGGTTTTTGATGTAGTACATGTTTCAAAAAGTACA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATAAAACGGAAAAAGCAAACACTAATCTTTTGGTTTTTGATGTAGTACATGTTTCAAAAAGTACA 540

Qy    541 TCCTTTTTAATCAGTTAACAATGTAGTTAGTGACCATGGTGTTAAATGGATTCCTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCCTTTTTAATCAGTTAACAATGTAGTTAGTGACCATGGTGTTAAATGGATTCCTT 600

Qy    601 TTGGGAATTCATGTATAAATTTACACATTACACATTCTTGTGATACTGAAATCTTTTTTTGCTGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTGGGAATTCATGTATAAATTTACACATTACACATTCTTGTGATACTGAAATCTTTTTTTGCTGA 660

Qy    661 GAAAGATTAAGTTGTCTTTGTTTGATTTTCATATAAAGCATCATGATGGTGTTTAATATTGT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAAAGATTAAGTTGTCTTTGTTTGATTTTCATATAAAGCATCATGATGGTGTTTAATATTGT 720

Qy    721 AAGATATTCTATAAGCAGTTGTGAAATCCAAAGTTCTCGTAAACATTTGTAGTGTTTG 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AAGATATTCTATAAGCAGTTGTGAAATCCAAAGTTCTCGTAAACATTTGTAGTGTTTG 780

Qy    781 AAATGAACAATGATATTATGAAGTGTGCTATCTGTAGACCTGACCTGGTGTAAGGACATTTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AAATGAACAATGATATTATGAAGTGTGCTATCTGTAGACCTGACCTGGTGTAAGGACATTTG 840
```

RESULT 4

```
LOCUS       AF031426                1040 bp    DNA     linear   ROD 25-DEC-1997
DEFINITION  Mus musculus small unique nuclear receptor co-repressor (SUN-CoR)
            gene, complete cds.
ACCESSION   AF031426
VERSION     AF031426.1  GI:2642585
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1040)
AUTHORS     Zamir,I., Dawson,J., Lavinsky,R.M., Glass,C.K., Rosenfeld,M.G. and
            Lazar,M.A.
TITLE       Cloning and characterization of a corepressor and potential
            component of the nuclear hormone receptor repression complex
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 94 (26), 14400-14405 (1997)
MEDLINE     98070763
PUBMED      9405624
REFERENCE   2  (bases 1 to 1040)
AUTHORS     Zamir,I. and Lazar,M.A.
TITLE       Direct Submission
JOURNAL     Submitted (24-OCT-1997) Medicine, University of Pennsylvania, 611
            CRB, 415 Curie Blvd., Philadelphia, PA 19104, USA
FEATURES             Location/Qualifiers
     source          1..1040
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            80..505
                     /gene="SUN-CoR"
                     /note="corepressor for nuclear hormone receptors; CID"
     CDS             80..505
                     /gene="SUN-CoR"
                     /codon_start=1
                     /product="small unique nuclear receptor co-repressor"
                     /protein_id="AAC53520.1"
                     /db_xref="GI:2642586"
                     /translation="MAGEEMNEDYPVEIHESLTALRESSLGAVDDMLKTMMAVSRNELL
                     KKKAAKLDRGASRFVKNALWEPKAKSTPKVANKGKSKH"
ORIGIN
```

Query Match   92.3%; Score 960.4; DB 10; Length 1040;
Best Local Similarity   97.9%; Pred. No. 2.4e-181;
Matches 1016; Conservative   0; Mismatches 16; Indels 6; Gaps 4;

```
Qy          1 CAGAAGCCGTGTCATGGGGTCATCATCGTTGCGACCTATTTCCCGGAGCAGACAGGCGTTCCACG  60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 CAGAAGCCGTGTCATGGGGTCATCATCGTTGCGACCTATTTCCCGGAGCAGACAGGCGTTCCACG  62

Qy         61 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAATTC  120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAATTC  120
```

```
Qy        841 TTTTTCAGTAAATGATGAGAGAAATACAGTGAACTTAAATACCACTCGTTCTCGTTCAGTTAG  900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTTTCAGTAAATGATGAGAGAAATACAGTGACTTAAATACCACTCGTTCTCGTTCAGTTAG  900

Qy        901 TTCAACATGTTTCGTGATTTTTTTTTGAGTAATACCACTTAAATACCAGTTCAAAGTC  960
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db        901 TTCAACATGTTTCGTGATTTTTTTTTTGAGTAATACCACTTAAATACCAGTTCAAAGTC  960

Qy        961 AAAAATTGAAAACCTTAAGGGCTGTACTTTAAATCTTCATGTTCCATTTAAATAAAATGTTC  1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAAATTGAAAACCTTAAGGGCTGTACTTTAAATCTTCATGTTCCATTTAAATAAAATGTTC  1020

Qy       1021 TCATTAACTCTGATGGAAAA  1040
              ||||||||||||||||||||
Db       1021 TCATTAACTCTGATGGAAAA  1040
```

RESULT 5

```
LOCUS       BC005436                2000 bp    mRNA    linear   ROD 29-JUN-2004
DEFINITION  Mus musculus nuclear DNA binding protein, mRNA (cDNA clone MGC:5983
            IMAGE:3591682), complete cds.
```

```
Db         63 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAATTC  122

Qy        121 ACGAGTCTTTAACAGCCCTGGGAGAGCTCCCTGGGGTCTCCCTGGGTGGTGAGCAGAGGCCA  180
Db        123 ACGAGTCTTTAACAGCCCTGGGAGAGCTCCCTGGGGTCTCCCTGGGTGGTGAGCAGAGGCCA  182

Qy        181 TGATGGCGTGTTTCTAGAAACGAGTTGTTGGCGAAGTTGGACCCATTGGGAACACAGCAAAGG  240
Db        183 TGATGGCGTGTTTCTAGAAACGAGTTGTTGGCGAAGTTGGACCCATTGGGAACACAGCAAAGG  242

Qy        241 TGGATTTTAGTTTTCTGCATCACCACCTTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAG  300
Db        243 TGGATTTTAGTTTTCTGCATCACCACCTTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAG  302

Qy        301 GAGTTAATCCCAAAAGAGACATCCAGTGAAGCAGGAAACTGGAAAGAAGAATCAGAGTCTACATGA  360
Db        303 GAGTTAATCCCAAAAGAGACATCCAGTGAAGCAGGAAACTGGAAAGAAGAATCAGAGTCTACATGA  362

Qy        361 ACAGAGTTAAAGAAAATAAACAACAAGAAGAGGGCTGCCAAGCTGCGACAGAGGTGCTGCTT  420
Db        363 ACAGAGTTAAAGAAAATAAACAACAAGAAGAGGGCTGCCAAGCTGCGACAGAGGTGCTGCTT  422

Qy        421 CGAGATTTGTCAAGAAGGCACTCTCGGGACACCCAAACAGAAAAAGCACACACAAAAGTGGCTA  480
Db        423 CGAGATTTGTCAAGAAGGCACTCTCGGGACACCCAAACAGAAAAAGCACACACAAAAGTGGCTA  482

Qy        481 ATAAAGGGGAAAAAGCAAACACTAATCTTTTTGGTTTTGATGTAACATGATGTTTTCAAAAAGTACA  540
Db        483 ATAAAGGGGAAAAAGCAAACACTAATCTTTTTGGTTTTGATGTAACATGATGTTTTCAAAAAGTACA  542

Qy        541 TCCTTTTTAATCAGTTACAATGTAGTTATTGTGATGTTGTTTAAAATGGATTCCTT  600
Db        543 TCCTTTTTAATCAGTTACAATGTAGTTATTGTGATGTTGTTTAAAATGGATTCCTT  602

Qy        601 TTGGAATTCATGTATAAATTTACACATTACACATTGTGATACTGAATCTTTTTTTTTGTGA  660
Db        603 TTGGAATTCATGTATAAATTTACACATTACACATTGTGATACTGAATCTTTTTTTTTGTGA  662

Qy        661 GAAAGATTAAGTTGTCTTTTTGTTGATTTCATTAAAAGCATCATGATGTGTTTAATATTGT  720
Db        663 GAAAGATTAAGTTGTCTTTTTGTTGATTTCATTAAAAGCATCATGATGTGTTTAATATTGT  722

Qy        721 AAGATATTCTATAAGCAGTTGTGAAAATCCAAATGTTCTCTGTAAACATTGTAGTGTTTG  780
Db        723 AAGATATTCTATAAGCAGTTGTGAAAATCCAAATGTTCTCTGTAAACATTGTAGTGTTTG  782

Qy        781 AAATGAACAATGATATTATGAAGTGTCCTACTCTGTAGACCTCGGAGGCTGTAAGGACATTG  840
Db        783 AAATGAACAATGATATTATGAAGTGTCCTACTCTGTAGACCTCGGAGGCTGTAAGGACATTG  841

Qy        841 TTTTCAGTTAATGATGAGAAATACAGTGACTTAAATACCCACTCGTTCTCGTTCAGTTA  899
Db        842 TTTTCAGTTAATGATGAGAAATACAGTGACTTAAATACCCACTCGTTCTCGTTCAGTTA  901

Qy        900 GTTCAACATGTTTCGTGATTTTTTTTGAG---TAATTCGTCTTGATATTCAA  956
Db        902 GTTCAACATGTTTCGTGATTTTTTTTTGAG---TAATTCGTCTTGATATTCAA  961

Qy        957 AGT--CAAAATTGAAACCTTAAGGCTGTACTTTAAATTCTTCATGTTCCATTTAAATAAAA  1015
Db        962 AGTACAAAATTGAAACCTTAAGGCTGTACTTTAAATTCTTCATGTTCCATTTAAATAAAA  1021

Qy       1016 TGTTCTCATTAACTCTGA  1033
Db       1022 TGTTCTCATTAACTCTGA  1039
```

```
ACCESSION   BC005436
VERSION     BC005436.1  GI:13529391
KEYWORDS    MGC.
SOURCE      Mus musculus
  ORGANISM  Mus musculus (house mouse)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2000)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2000)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil1.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 11 Row: d Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 10181123.
FEATURES             Location/Qualifiers
     source          1..2000
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C3ECH II"
                     /db_xref="taxon:10090"
                     /clone="MGC:5983 IMAGE:3591682"
                     /tissue_type="Mammary tumor metastasized to lung. Tumor
                     arose spontaneously from a senescent normal mammary
                     /clone_lib="NCI_CGAP_Lu29"
                     /lab_host="DH10B"
                     /clonal) outgrowth infected with the virus MMTV."
                     /note="Vector: pCMV-SPORT6"
     gene            1..200
                     /gene="Cld"
                     /note="synonym: SUN-COR"
                     /db_xref="LocusID:57316"
                     /db_xref="MGI:1927354"
     CDS             47..472
                     /gene="Cld"
                     /codon_start=1
                     /product="nuclear DNA-binding protein"
                     /protein_id="AAH05436.1"
                     /db_xref="GI:13529392"
                     /db_xref="LocusID:57316"
                     /db_xref="MGI:1927354"
                     /translation="MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVYTLATQVGVNPKEHPVKQELERIRIVYNRVKEITD
                     KKKAAKLDRGAASRFVKNALWEPKAKSTPKVANKGKSKH"
ORIGIN
            Query Match         90.8%;  Score 944.6;  DB 10;  Length 2000;
            Best Local Similarity   97.8%;  Pred. No. 3.1e-178;
            Matches  989;  Conservative    0;  Mismatches   19;  Indels    3;  Gaps    3;
```

```
Qy         32 GACCTATTTCCCGGAGACAGGCGGTCCACGGTATTGAGTTGGTCACAATGGCAGGTGAAGA         91
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACCTATTTCCCGGAGACAGGCGGTCCACGGTATTGAGTTGGTCACAATGGCAGGTGAAGA         60

Qy         92 AATGAATGAAGATTATCCCGTAGAAAATTCACGAGTCTTTAACAGCCCCTGGAGAGCTCCCT        151
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AATGAATGAAGATTATCCCGTAGAAAATTCACGAGTCTTTAACAGCCCCTGGAGAGCTCCCT        120

Qy        152 GGGGTGCTGTGGACGACATGCTGAAGACCATGATGGCTGTTCTAGAAACGAGTTGTTGCA        211
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGGGTGCTGTGGATGACATGCTGAAGACCATGATGGCTGTTCTAGAAATGAGTTGTTGCA        180

Qy        212 GAAGTTGGACCCCATTGGAACAAAGCAAAAGGTGGATTTAGTTTCTGCATACACCTTAAATTC        271
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAGTTGGACCCCATTGGAACAAAGCAAAAGGTGGATTTAGTTTCTGCATACACCTTAAATTC        240

Qy        272 AATGTTTTGGGTTTATTTGGCAACTCAACGAGTTAATCCCAAAGAGCATCCAGTGAAGCA        331
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AATGTTTTGGGTTTATTTGGCAACTCAACGAGTTAATCCCAAAGAGCATCCAGTGAAGCA        300

Qy        332 GGAACTGGAAAGAATCAGAGTCTACATGAACAGAGTTAAAGAAATAACGACACAAGAAGA        391
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAACTGGAAAGAATCAGAGTCTACATGAACAGAGTTAAAGAAATAACGACACAAGAAGA        360

Qy        392 GGCTGCCAAGCTGGACAGAGGTGCTGCTTCGGATTGTGCAAGAAGGCACACTCTGGAAACC        451
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCTGCCAAGCTGGACAGAGGTGCTGCTTCGGATTGTGCAAGAACGGCACACTCTGGAAACC        420

Qy        452 CAAACGAAAAAAGCACACCAAAAGTGGCTAATAAAAGTGGCAAAAGCAAAACACTAATCTTTTGG        511
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAAACGAAAAAAGCACACCAAAAGTGGCTAATAAAAGGGAAAAAGCAAAACACTAATCTTTTGG        480

Qy        512 TTTTGAATGTACATGTTTTCAAAAAGTACATCCTTTTTTAAATCAGTTTGATTTTTCAT        571
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTTTGAATGTACATGTTTTCAAAAAGTACATCCTTTTTTAAATCAGTTTGATTTTTCAT        540

Qy        572 TGACCATGTGGTGTTTAAATGGATTCCTTTGGAATTCATGTATAAAATTTACACATTACA        631
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGACCATGTGGTGTTTAAATGGATTCCTTTGGAATTCATGTATAAAATTTACACATTACA        600

Qy        632 TTTGTGATACTGGAATCTTTTTTTGCTGAAAAGAATTAAGTGTCTTGTTGTTGATTTTCAT        691
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTTGTGATACTGAATCTTTTTTTTGCTGAAAAGAATTAAGTGTCTTGTTGTTGATTTCAT        660

Qy        692 ATAAAAGCCATCATGATGTGTTAAATTGTAAGATATTCTAATAAGCAGTTGTTGAAATCCAA        751
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ATAAAAGCCATCATGATGTGTTTAAATTGTAAGATATTCTAATAAGCAGTTGTTGAAATCCAA        720

Qy        752 ATGTTCTCTGTAAAACATTTGTAGTGTTTGTTGAAATGAACAATGATATTATGAAGTGTCCTAT        811
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATGTTCTCTGTAAAACATTTGTAGTGTTTGTTGAAATGAACAATGATATTATGAAGTGTCCTAT        780

Qy        812 CTGTAGGACCTCGAGGTGTAAGGACCATTGTTTTTTCAGTAATGATGAGAAAATACAGTGACTT        871
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGTAGA-CTGGAGGTGTAAGGACCATTGTTTTTTCAGTAATGATGAGAAAATACAGTGACTT        839
```

```
Qy   872  AAATACCCACTCTG--TTTCTGTTGTTCAGTTAGTTCAACATGTTTCGTGGATTTTTTTTTTTTTT
          |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db   840  AAATACCCACTCTGTTTTCTGTTGTTCAGTTAGTTCAACATGTTTCGTGGATTTTTTTTTTTTT  899

Qy   931  TTGAGTAATTCTGTCTTTGATATTCAAAGT-CAAAATTGAAACCTTAAGGCTGTACTTTAA  989
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   900  GGAGTAATTGCTGTCTTTGATATTCAAAGTACAAAATTGAAACCTTAAGGCTGTACTTTAA  959

Qy   990  TTCTTCATGTTCCATTTAAATAAAATAGTTCTCATTAACTCTGATGGAAAA  1040
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   960  TTCTTCATGTTCCATTTAAAATAAAATGTTCTCATTAACTCTGATGGAAAA  1010
```

RESULT 6
ACI09025/c

LOCUS   ACI09025
DEFINITION Rattus norvegicus clone CH230-276A24, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   ACI09025
VERSION   ACI09025.5 GI:25139227
KEYWORDS   HTG; HTGS_PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 215616)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,M., Hamill,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,J., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarinpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,J., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Shen,H.,
            Sneed,N., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingley,A., Trejos,Z., Usmani,K.,
            Valas,V., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wlczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 215616)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 215616)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE   Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Nov 20, 2002 this sequence version replaced gi:23603213.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contig will be indicated in the feature
            table.
            -------------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------------- Project Information
               Center project name: GPRH
               Center clone name: CH230-276A24
            -------------------- Summary Statistics
               Assembly program: Phrap; version 0.990329
               Consensus quality: 181529 bases at least Q40
               Consensus quality: 183677 bases at least Q30
               Consensus quality: 185417 bases at least Q20
               Estimated insert size: 187302; sum-of-contigs estimation
               Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            ----
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 1 contigs. Gaps between the contigs
            *   are represented as runs of N. The order of the pieces
            *   is believed to be correct as given, however the sizes
            *   of the gaps between them are based on estimates that have
            *   provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *
FEATURES   Location/Qualifiers
     source   1..215616
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-276A24"
     misc_feature   1..1067
                     /note="wgs_end_extension
     misc_feature   clone_end:Sp6"
                     8625..10254
                     /note="wgs_end_extension
     misc_feature   clone_end:Sp6"
                     11405..12230
                     /note="clone_boundary
                     clone_end:Sp6"
```

```
Db    misc_feature    site:
                      end_sequence:BZ161264"
                      14323..150442
      misc_feature    /note="wgs_contig"
                      complement(212205..213062)
Qy                    site:
                      clone_end:T7
Db                    /note="clone_boundary"
                      end_sequence:BZ161263"

ORIGIN

Qy    Query Match    73.8%;   Score 767.8;   DB 2;   Length 215616;
Db    Best Local Similarity  89.0%;   Pred No. 3.2e-143;
      Matches 932;  Conservative  0;  Mismatches 102;  Indels 13;  Gaps  9;

Qy         1 CAGAAGCCCTGTCATCGGCGTCATCATCGTCGTGGCGGACCTTATTTCCCGGAGACGAGGGGGTCCACG    60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181953 CAAAAAACCCGTGTCACAAATGGCGGTCATCATCTGCCAGGCAGGCGCGTCCACG    181894

Qy        61 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAGAGATTATCCGGTAGAAATTC    120
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db    181893 GTATTGAGTTGGTCACAATGGCAGGTGAAGAAATGAAGAGATTATCCGGTAGAAATTC    181834

Qy       121 ACGAGTCTTTAACAGCCCTGGGAGACTCCCTGGGTGTCTGTGGGACACATGCTGAGACCA    180
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181833 ACGAGTTCTTTGACAGCTCTGGAGAGCTCCCTGGGGCGGTGGATGACATGCTGAGACCA    181774

Qy       181 TGATGGCTGCTTTCTTAGAAACGAGTGTTGTTGCAGAAAGCAAGCAAGCAAAAGG    240
             |||||||||||||||||||||||||||||||||||||||||||||||
Db    181773 TGATGTCTGTTCTTCTAGAAATGAGTTGTTGTTGCAGAAGTGGACCCACTGGAACAAGGAAAGG    181714

Qy       241 TGGATTTAGTTTCTGCATACACCTTAAATTCAAATGTTTTGGGTTTATTTGGCAACTCAAG    300
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db    181713 TGGATTTAGTTTCTGCATATACATTAAATTCAAATGTTTTACTTGGCAACTCAAG    181654

Qy       301 GAGGTCAAAAGAGCACCCCAGTGAAGCGGGAAGAATTCAGAGTCTTACATGA    360
             |||||||||||||||||||||||||||||||||||||||||||||||
Db    181593 ACAGTCAAAAGAATAACAGACAAGAAGGCTGCCAAGCTCAGAGGGCCCGGT    181534

Qy       361 ACAGAGTTAAAGAAATAACAGAGAAAGAGGCTGCCAAGCTGGAACAAGCTGCTGCTT    420
             ||||||||||||||||||||||||||||||||||||||||||||||
Db    181653 GAGTTAAATCCCAAGGACATCCAGTGGAAGAGGGAGAATCCGAGTGTTCACATGA    181594

Qy       421 CGAGATTTGTCAGNAGGCACTCTGGGGAACCCAAACAAAAAAGCACACACAAAAGTGGCTA    480
             ||||||||||||||||||||||||||||||||||||||||||||||
Db    181533 CGAGATTTGTCAAGAATTCGACTCTCGGGGAACCCAAAACAAAAAAGTGGGCCGCGT    181474

Qy       481 ATAAAGGGAAAAGCAAACAC---TAATCTTTGGTTTTGATTGTACATGTTTTTCAAAAAGT    537
             ||||||||||||||||||||||| ||||||||||||||||||||||
Db    181473 ATAAAGGGAAAAGCTAAACACTAATAATCTTTTGGTTTTGATGTACGTGTTTTCAAAAAGT    181414

Qy       538 ACATCCCTTTTAATCCAGTTTACAAGTAGTGTTGACCATGGTGTGTTTAAATGGATTC    597
             ||||||||||||||||||||||||||||||||||||||||||
Db    181413 ACATCATTTTT-ATCAGTTTACAGTGTACCATGTCGTGTTTAAATGGATTC    181355

Qy       598 CTTTTGGAAGTTCATGTATATAAAATTACACATTACA-TTTGTGATACTGAATC-TTTTTTTT    655
             ||||||||||||||||||||||||||||||||||| |||||||||| |||||||
Db    181354 CTTTTGAATTCATCGTGATATAAAATTTGTGATACTGAATCTTTTTTTT    181295

Qy       656 GCTGAGAAAGATTAAGTTGTCTTTGTCTTTTTGATTTTCATATAAAGCATCATGATTGTTTTAAT    715
             ||||||||||||||||||||||||||||||||||||||||||||||
Db    181294 GCTGAGAAAGATTAAGTTGTTCGTCATTTTCTTTTTGATTTCATATAAAGCATCATGATTGT-TTTAAT    181236

Qy       716 ATTGTTAAGATATTCTATAAGGCAGTTGTGAAATCCAAATGTTCTCTGTAAACATTTGTAGT    775
             ||||||||||||||||||||||||||||||||||||||||||||||
Db    181235 ATTGAAAAGATATTCTATAAGCAATTGTGAAATGTTCTTTGTAAACGTTTGTAGT    181176

Qy       776 GTTTGAAATGAACAAATGATATTATGAAGTGTGCTARTCTGTAGACCTTCGAGGTGTAAGGAC    835
             ||||||||||||||||||||||||||||||||||||||||| ||
Db    181175 GTTTTAAATG--TAGTGTATATGAGTGAAAGTGTCGTCGTTAGTCTGTAGAGGA    181118

Qy       836 ATTGTGTTTCAGTAATGATGAGAAATACAGTGACTTAAAATACCCACTCT-GTTTCTGTGTTC    894
             ||||                 |||||||
```

```
Db    181117 CATCTGTTTTCACATGTGAGAAAATGTAATGAGAAGTTAAATACCCACTCTGGTTTCTGTTC    181058
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       895 AGTTAGTTCAACATGTTTCGTGCATTTTTTTTTTTTTGAGTAATTGCTTTCTGTCTGTGATATTC    954
             |||||||||||||||||| |||||||||||||||||| ||| |||||||
Db    181057 AGTTAGTTCAACATGTTTC--TGTGAATTTTTTTAGGAGTAATTGCTGTCTTGAAGTCCTTCATTTAAATAA    181000

Qy       955 AAAGT-CAAAAATTGAAACCTTAAGGCTGTACTTTAATTCTTCATTGTTCCATTTAAATAA    1013
             ||||| |||||||||||||||||||||||||||||||||||||
Db    180999 ACAGTACAAAAATTGAAACCTTCAGGCGGTGTACTTTAAGTTCTTCATTGTTCCATTTAAATAA    180940

Qy      1014 AATGTTCTCATTAACTCTGATGGAAA    1040
             |||||||||||||||||||| |||
Db    180939 AATATTCTCATTCATTAAAAAAAAAAA    180913
```

```
TITLE             Direct Submission
JOURNAL           Unpublished
REFERENCE         2  (bases 1 to 224919)
AUTHORS           Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
                  Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
                  Niederhausern,A., Weiss,R., Smith,D.R., Dunn,D., von
                  Weinstock,G. and Gibbs,R.A., Smith,D.R., Holt,R.A., Smith,H.O.,
                  Worley,K.C.
TITLE             Direct Submission
JOURNAL           Submitted (11-FEB-2002) Human Genome Sequencing Center, Department
                  of Molecular and Human Genetics, Baylor College of Medicine, One
                  Baylor Plaza, Houston, TX 77030, USA
REFERENCE         3  (bases 1 to 224919)
AUTHORS           Rat Genome Sequencing Consortium.
TITLE             Direct Submission
JOURNAL           Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
                  of Molecular and Human Genetics, Baylor College of Medicine, One
                  Baylor Plaza, Houston, TX 77030, USA
COMMENT           On Nov 15, 2002 this sequence version replaced gi:23603309.
                  The sequence in this assembly is a combination of BAC based reads
                  and whole genome shotgun sequencing reads assembled using Atlas
                  (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
                  in the feature table below represents a scaffold in the Atlas
                  assembly (a 'contig-scaffold'). Within each contig-scaffold,
                  individual sequence contigs are ordered and oriented, and separated
                  by sized gaps filled with Ns to the estimated size. The sequence
                  contigs within a contig-scaffold that consist entirely of whole
                  shotgun sequence reads. Both end sequences and whole genome
                  genome shotgun sequence only contigs will be indicated in the feature
                  table.
                  -------------- Genome Center
                  Center: Baylor College of Medicine
                  Center code: BCM
                  Web site: http://www.hgsc.bcm.tmc.edu/
                  Contact: hgsc-help@bcm.tmc.edu
                  -------------- Project Information
                  Center project name: GOEH
                  Center clone name: CH230-224B18
                  -------------- Summary Statistics
                  Assembly program: Phrap; Version 0.990329
                  Consensus quality: 215079 bases at least Q40
                  Consensus quality: 216539 bases at least Q30
                  Consensus quality: 217655 bases at least Q20
                  Estimated insert size: 221721; sum-of-contigs estimation
                  Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
                  --------------
                  * NOTE: Estimated insert size may differ from sequence length
                  *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
                  * NOTE: This is a 'working draft' sequence. It currently
                  * consists of 1 contigs. Gaps between the contigs
                  * are represented as runs of N. The order of the pieces
                  * is believed to be correct as given, however the sizes
                  * of the gaps between them are based on estimates that have
                  * provided by the submitter.
                  * This sequence will be replaced
                  * by the finished sequence as soon as it is available and
                  * the accession number will be preserved.
                  *
FEATURES             Location/Qualifiers
     source          1..224919
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-224B18"
     misc_feature    1..1155
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    2063..2995
                     /note="clone_boundary
                     clone_end:Sp6"
                     site:
                     end_sequence:BZ099339"
     misc_feature    complement(221675..222607)
                     /note="clone_boundary
                     clone_end:T7"
                     site:
                     end_sequence:BZ099339"
     misc_feature    223559..224919
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN
     Query Match
     Best Local Similarity   89.0%;
     Matches  932;  Conservative    0;  Mismatches  102;  Indels  13;  Gaps  9;
```

```
Qy          1 CAGAAGACGCGTGCGTGTCATGCCGTCCATCATCGTGCGGGGACCTAATTCCCGGAGACAGCAGGGCGTCCACG         60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186006 CAAAAACCGTGTGCATGCCGTCATCATTGCGCGCACCTCTTTTCCCGGAGACAGCAGGGCGTCCACG         186065

Qy         61 GTATTGAGTTGGTCTCACAAATGGCCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAAATTC        120
              ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     186066 GTAATTGAGTTGGTCACAAATGGCCAGGTGAAGAAATGAATGAAGACTAATCCCGTAGAAAATTC        186125

Qy        121 ACGAGTCTTTAACAGCCCTGGAGAGCTCCCTGGGTTGCTGTGGGACGACAACATGCTGAAGACCA        180
              |||||||||||||||||||| ||||||||||||| ||||| ||||||||||||||||||||||||
Db     186126 ACGATTCTTTGACAGCTCTCGGAGACTCCCTGGGGCCGTGGATGACAACATGCTGAAGACCA        186185

Qy        181 TGAATGGCTCGTTTCTAGAAACGAGGTTGTTCGGAGAAGTTGGACCCACTGGAACAAGCAAAGG        240
              ||||||||||| ||||||||||||||||| ||||| |||||||||||||||||||||||||||||
Db     186186 TGATGTCTCGTTTCTAGAAATGAGTTGTTCGGAGAAGTTGGACCCACTGGAACAAGCAAAGG        186245

Qy        241 TGGATTTAGTTTTCTGCATACACTTTAAATTCAAATGTTGTTTTGGTTTATTGGGCGAACTCAAG        300
              ||||||||||||| ||||||||||||||||||||||| |||||||| ||||||||||||||||||
Db     186246 TGGATTTAGTTTTCTGCATATACATTACAAATTGTTTGTTACTTGGGGCGAACTCAAG        186305

Qy        301 GAGTTAAATCCCAAAGAACAGCATCCAGTGAAGCAGGAACTGAAGAGAAATCAGAGTCTACATGA        360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186306 GAGTTAAATCCCAAAGGAGCAGCATCCAGTGAAGCAGGAACTGAAGAGAAATCGAGTCTACATGA        186365

Qy        361 ACAGAGTTAAAGAATAACAGAGAAGAGAAGGAGGCTGCCAAGCTGGACGAGAGGTGTCGCTT        420
              |||||||||||||||||||||||||||||||||||||||||||| || |||| ||||| |||
Db     186366 ACAGAGTCAAAGAAATAACAGAGAAGAGGCTGCCCAAGCTGGACAGCAGGGGGCGCGT        186425

Qy        421 CCGAGATTTGTCAAGAAGGCACTCTCGGAGAACCCAACCGAAAAAAGCACACCAAAAGTGGCTA        480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186426 CCGAGATTTGTCAAGAAGGCACTCTCGGGACCCAACCGAAAACAACAACAAAAGTGGCTA        186485

Qy        481 ATAAAGGGAAAAGCAAACAC-----TAATCTTTTGGTTTTGGTTTGATGTATGTTTCAAAAAGT        537
              ||||||| ||||||||||||     |||||||| |||||| |||||||||| ||||||||| ||
Db     186486 ATAAAGGGAAAAGTAAACACTAATAATCTTTTGGATTGTTTGATGTGTATGTTTCAAAAAGT        186545

Qy        538 ACATCCTTTTAATCAGTTTACAATGTAGTTATGTGACCATGTGGTGTTTAAATGGATTC        597
              |||||||||||||||||||||||| ||||| ||||| ||||||||||||||||||| |||
Db     186546 ACATCATTTT-ATCAGTTTACAGTTGTACTTATGTGACCATGTGGTGTTTAAATGGATTC        186604

Qy        598 CTTTTGGAATTCATGTATAATTTACACATTACA-TTTGTGATACTGAATC-TTTTTTT        655
              ||||||||||||||||||||||||||||||||| ||||||| ||| |||| ||||||||
Db     186605 CTTTTTGGAATTCATGTATAAAATTTACACATTAAATTTGTGATACTGAATCTTTTTTTT        186664

Qy        656 GCTGAGAAAGATTAAGTTGCTCTTTGTTGATTTTGATTTCATATAAAGGCATCATGAGTGTTTAAT        715
              |||||||||||||||||| |||||||| ||||||||||||||||||||||||| ||||||| ||||
Db     186665 GCTGAGAAAGATTAAGTTGTTCGTTGTTATTGATTTTCATATAAAGGCATCATGAGTGTTAAT        186723

Qy        716 ATTGTAAGATATTCTTATAAGCACAGTTGTGAAAATCCAAAATGTTCTCTGTAAACATTGTAGT        775
              ||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     186724 ATTGAAAGATATTCTATAAGCAATGTGAAATCCAAAATGTTCTTGTAAAACGTTGTGTAGT        186783

Qy        776 GTTTGAAATGAACAAATGATTATTATGAAGTGTGCTATCTGTAGACCTCGAGGTGTAAGGAC        835
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     186784 GTTTTAAATG--TAGTGTTTATGATGAAGTGTGCGTTCTGTAGAGGTGTTAGAGGA        186841

Qy        836 ATTTGTTTTCAGTAATGATGAGAAATACAGTGACTTAAATACCCACTCT-GTTTCTGTGTTC        894
              ||||||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||||
Db     186842 CACCTGTTTTCACAATGATGAGAAAGTGAATGACTTAAATACCCACTCTGGTTTCTGTGTTC        186901
```

```
Qy    895 AGTTAGTTCAACACATGTTTCGTGTGATTTTTTTTTTTTGAGTAATTCTGTGCTTGATATTC 954
          |||||||||||||||||||||||||||||||||||    ||||||| ||||||||||| |||
Db 186902 AGTTAGTTCAACACATGTTC--TGTGAATTTTTTTTTAGGAGTAATTGCTGTGCTTGATGTTC 186959

Qy    955 AAAGT-CAAAATTGAAACCTTAAGGCTGTACTTTAATTCTTCATGTTCCATTAAAATAA 1013
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 186960 ACAGTACAAAATTGAAACCTTCAGGCCGTACTTTAATTCTTCATGTTCCATTAAAATAA 187019

Qy   1014 AATTGTTCTCATTACTCTGATGGAAA 1040
          ||||||||||||||||||||||||||
Db 187020 AATATTCTCATTCATTAAAAAAAAAA 187046
```

RESULT 8

LOCUS   AL603925   232869 bp   DNA   linear   ROD 22-MAY-2002
DEFINITION Mouse DNA sequence from clone RP23-175K24 on chromosome 11,
            complete sequence.
ACCESSION   AL603925
VERSION   AL603925.20 GI:21211846
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 232869)
AUTHORS   Bird,C.
TITLE   Direct Submission
JOURNAL   Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
            humquery@sanger.ac.uk
COMMENT   On May 25, 2002 this sequence version replaced gi:20803491.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30).; an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-175K24 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES   Location/Qualifiers
     source   1..232869
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-175K24"
                     /clone_lib="RPCI-23"
ORIGIN
Query Match   61.7%; Score 642; DB 10; Length 232869;
Best Local Similarity   98.0%; Pred. No. 3.8e-118;
Matches 692; Conservative   0; Mismatches 10; Indels 4; Gaps 4;

```
Qy    338 GGAAAGAATCAGAGTCTACATGAACAGAGAGAAATAACACAAGAAGAAGAGAGGCTGC 397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24099 GGAAAGAATCAGAGTCTACATGAACAGAGAGAAATAACACAAGAAGAAGAGAGGCTGC 24158
```

```
Qy    398 CAAGCTGGACAGAGGGTCGCTGCTTCGTGAGATTTGTTGTCAGAGAAGGCCACTCTGGGAACCCAAACG 457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24159 CAAGCTGGACAGAGGGTCGCTGCTTCGTGAGATTTGTTGTCAGAGAAGGCCACTCTGGGAACCCAAAGC 24218

Qy    458 AAAAAGCACACCCAAAAGTGGCTAATAAGGGGGAAAGGGGGAAAAGCAAAAACTAATACTTTTGGGTTTGA 517
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24219 AAAAAGCACACCCAAAAGTGGCTAATAAGTGGCTAATAAAGGGGGAAAAGCAAAAACTAATACTTTTGGGTTTGA 24278

Qy    518 TGTACATGTTTTCAAAAAGTACAATCCTTTTTAATCAGTTTACAAGTGACTTACAATGTAGTGTAGTGTGACCA 577
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24279 TGTACATGTTTCAAAAAGTACAATCCTTTTTAATCAGTTTACAAGTGACTTACAATGTAGTGTAGTGTGACCA 24338

Qy    578 TGTGGGTGTTTAAATGGATTCCTTTTGGAAGTTCATGTGTATATAAAATTTTACACATTTGTG 637
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24339 TGTGGGTGTTTAAATGGATTCCTTTTGGAAGTTCATGTGTATATAAAATTTTACACATTTGTG 24398

Qy    638 ATACTGAAATCTTTTTTTTTTTGCTGAGAAAGAGATTAAGTTCATGTTAGTATTCATATAAAG 697
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24399 ATACTGAAATCTTTTTTTTTTTGCTGAGAAAGAGATTAAGTTCATGTTAGTATTCATATAAAG 24458

Qy    698 CATCATCATGTGTCTTTAATATGTTTAAATATTGTTAAGATATTCTATAAGCAGTTGTGAAATCCAAATGTTC 757
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24459 CATCATCATGTGTCTTTAATATGTTTAAATATTGTTAAGATATTCTATAAGCAGTTGTGAAATCCAAATGTTC 24518

Qy    758 TCTGTAAACATTTGTTGTAGTGTTTGAAATGACAAATGATATTATGAAGTGTGTCTATCGTAG 817
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24519 TCTGTAAACATTTGTTGTAGTGTTTGAAATGACAAATGATATTATGAAGTGTGTCTATCGTAG 24578

Qy    818 ACCTCGAGGTGTAAGGACACATTGTTTGTTTCAGTAATGACAACAGTGACTTAAATAC 877
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24579 A-CTGGAGGTGTAAGGACACATTGTTTGTTTCAGTAATGAGAAATACAGTGACTTAAATAC 24637

Qy    878 CCACTCTG-TTTCTGTTTCAGTTAGTTCGTTGATTTTTTTTTTTTTTTTCAGT 936
          |||||||| |||||||||||||||||||||||||||||||||||||| ||||
Db  24638 CCACTCTGTTTTCTGTTTCAGTTAGTTCAACATGTTTCTGTGCAGTTTTTTTTTTGGAGT 24697

Qy    937 AATT--CTGTCTTGATATTCAAAGT-CAAAATTGAAACCTTAAGGCTGTACTTTAATTCTT 994
          ||||  ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  24698 AATTGCTGTCTTGATATTCAAAGTACAAAATTGAAACCTTCAGGCCGTACTTTAATTCTT 24757

Qy    995 CATGTTCCATTAAAATAAATAAAATGTTCTCATTAAACTCTGATGGAAA 1040
          |||||||||||||||||||     ||||||||||||||||||||||||
Db  24758 CATGTTCCATTAAAATAAAATGTTCTCATTAAACTCTGATGGAAA 24803
```

RESULT 9

LOCUS   AC044845   235365 bp   DNA   linear   HTG 23-MAR-2003
DEFINITION Mus musculus chromosome 11 clone RP23-175K24 map 11, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC044845
VERSION   AC044845.4 GI:29164568
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 235365)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE   Mus musculus chromosome 11 clone RP23-175K24
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 235365)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
```

AUTHORS   McCarthy,M., McEwan,P., McKernan,K., McPheeters,R.,
           Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
           Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
           Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
           Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
           Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
           Tesfaye,S., Theodore,J., Tirelli,A., Travers,M., Trigilio,J.,
           Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
           Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 235365)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
           Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
           Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
           Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
           Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
           Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
           Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
           Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
           Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
           Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
           Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
           Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
           Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
           Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
           Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
           Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
           Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 23, 2003 this sequence version replaced gi:1312929.
           All repeats were identified using RepeatMasker:
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           --------------------- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           --------------------- Project Information
           Center project name: L9172
           Center clone name: 175_K24
           --------------------- Summary Statistics
           Sequencing vector: M13; M77815; 32% of reads
           Sequencing vector: Plasmid; n/a; 68% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.960731
           Consensus quality: 233275 bases at least Q40
           Consensus quality: 233783 bases at least Q30
           Consensus quality: 234252 bases at least Q20
           Insert size: 220000; agarose-fp
           Insert size: 234765; sum-of-contigs
           Quality coverage: 13.9 in Q20 bases; agarose-fp
           Quality coverage: 13.0 in Q20.
           NOTE:This is a 'working draft' sequence. It currently
           consists of 7 contigs. The true order of the pieces
           is not known and their order in this sequence record is
           arbitrary. Gaps between the contigs are represented as
           runs of N, but the exact sizes of the gaps are unknown.
           This record will be updated with the finished sequence
           as soon as it is available and the accession number will
           be preserved.
           *   1
           *   9033: contig of 9033 bp in length
           *   9034: gap of 100 bp
           *   9034
           *   9133: gap of 100 bp
           *   9134: contig of 23820 bp in length
           *   32953: gap of 100 bp
           *   32954
           *   33053: gap of 100 bp
FEATURES   Location/Qualifiers
     source   1..235365
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /map="11"
                     /clone_lib="RPCI-23 Female Mouse BAC"
                     /clone="RP23-175K24"
     misc_feature   1..9033
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature   9134..32953
                     /note="assembly_fragment"
     misc_feature   33054..87658
                     /note="assembly_fragment"
     misc_feature   87759..117237
                     /note="assembly_fragment"
     misc_feature   117338..155012
                     /note="assembly_fragment"
     misc_feature   155113..207954
                     /note="assembly_fragment"
     misc_feature   208055..235365
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     *   33054   87658: contig of 54605 bp in length
     *   87659   87758: gap of 100 bp
     *   87759   117237: contig of 29479 bp in length
     *   117238   117337: gap of 100 bp
     *   117338   155012: contig of 37675 bp in length
     *   155013   155112: gap of 100 bp
     *   155113   207954: contig of 52842 bp in length
     *   207955   208054: gap of 100 bp
     *   208055   235365: contig of 27311 bp in length.

ORIGIN
     Query Match   61.7%; Score 642; DB 2; Length 235365;
     Best Local Similarity   98.0%; Pred. No. 3.8e-118;
     Matches 692; Conservative 0; Mismatches 10; Indels 4; Gaps 4;

Qy   338 GGAAAGAAATCAGAGGTCTACAATGAACAGAGTTAAAGAAAATAACAGAGCAAGAGAGCTGC 397
               |||||||||||||||||||||||||||||||||| || |||||||||||| |||||||||
Db   211267 GGAAAGAAATCAGAGTCTACAATGAACAGAGTTAAAGAAAATAACAGACAAGAAGAAGGCTGC 211208

Qy   398 CAAGCTGGACAGAGGGTGCTCGCTTCGAGAAAGATTAAGGTGGCTAATAAAAGGGAAAAAGC 457
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211207 CAAGCTGGACAGAGGGTGCTCGCTTCGAGAAAGATTAAGGTGGCTAATAAAAGGGAAAAAGC 211148

Qy   458 AAAAAGCACACCAAAAGTGGCTAATAAAGGGAAAAGCAAACACTAATCTTTTGGTTTTGA 517
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211147 AAAAAGCACACCAAAAGTGGCTAATAAAAGGGAAAAGCAAACCCACTCTGGGGAACCCAAAGC 211088

Qy   518 TGTACAATGTTTTCAAAAAAGTACATCCTTTTTAATCAGTTTACAAATGGTAGTGTGACCA 577
               |||||||||||||||||||| |||||||||||||||||||||||||||| |||||||||||
Db   211087 TGTACAATGTTTTCAAAAAGTACATCCTTTTTAATCAGTTTACAAATGGTAGTGTGACCA 211028

Qy   578 TGTGGGTGTTTAAATGGATTCCTTTTTTGGAAATTCATTGTATAAAATTTACAATTACATTGTG 637
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211027 TGTGGGTGTTTAAATGGATTCCTTTTTTGGAAATTCATTGTATAAAATTTACACATTACATTGTG 210968

Qy   638 ATACTGAATCTTTTTTTTTGCTGAGAAGAGATTAAGTTGTCTTGTTGATTTTCATATAAAG 697
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210967 ATACTGAATCTTTTTTTTTGCTGAGAAGAGATTAAGTTGTCTTGTTGATTTTCATATAAAG 210908

Qy   698 CATCATGATGTGTTTAATAATGTGTAGCATATTCTATAAGCAGATATTCTATAAGCAACAATGTTC 757
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210907 CATCATGATGTGTTTAATAATGTGTAAGATATTCTATAAGCAGTGTGAAATCCAAAATGTTC 210848

Qy   758 TCTGTAAACATTTGTAGTGTTTGAAATGAACAATGATATTCATATGCTATTCGTAG 817
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210847 TCTGTAAACATTTGTAGTGTTTGAAATGAACAATGATATTCATATGAGTGTGCTATTCGTAG 210788

```
Qy     818 ACCTTCGAGGTGTAAGGACATTTGTTTTCAGTAATGAGTAGTGATGAGAAATACAGTGACTTAAATAC 877
               |||||||||||||||||||||||||||||||||||||||||||||||||
Db  210787 A-CTGGAGGTGTAAGGACAATTGTTTTCAGTAATGATGAGAAATACAGTGACTTAAATAC 210729

Qy     878 CCACTCTG-TTTCTGTTCAGTTAGTTCAACATGTTTCGTCGATTTTTTTTTTTTTTGAGT 936
               ||||||||     ||||||||||||||||||||||||||||||||||||||||||
Db     878 CCACTCTGTTTTCTGTTCAGTTAGTTCAACATGTTTCGTCGATTTTTTTTTTTTTTGGAGT 936

Qy  210728 CCACTCTGTTTTCTGTTCAGTTAGTTCAACATGTTTCGTCGATTTTTTTTTTTTGGAGT 210669

Qy     937 AATT-CTGTCTTGTTGATTATTCAAAGT-CAAAATTGAAACCTTAAGGCTGTACTTTAATTCTT 994
               ||||  ||||||||||||||||||||    |||||||||||||||||||||||||||||
Db  210668 AATTGCTGTCTTGATATTCAAAGTACAAAGTTAAGGCTGTACTTTAATTCTT 210609

Qy     995 CATGTTCCATTAAAATAAAAATAAAATGTTCTCATTAACTCTGATGGAAAA 1040
               ||||||||||||||||||||||||||||||||||||||||||||||
Db  210608 CATGTTCCATTAAAATAAAAATGTTCTCATTAACTCTGATGGAAAA 210563
```

RESULT 10

LOCUS   AY302220   789 bp   mRNA   linear   ROD 28-MAY-2003
DEFINITION   Cricetulus griseus nuclear DNA-binding protein CID mRNA, complete
               cds.
ACCESSION   AY302220
VERSION   AY302220.1 GI:31095451
KEYWORDS   .
SOURCE   Cricetulus griseus (Chinese hamster)
  ORGANISM   Cricetulus griseus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
               Cricetulus.
REFERENCE   1 (bases 1 to 789)
  AUTHORS   Guang,L., Masabumi,S. and Maru,Y.
  TITLE   Differential display analysis of BCR-ABL-regulated genes
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 789)
  AUTHORS   Guang,L., Masabumi,S. and Maru,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-MAY-2003) Dept. Genetics, The Institute of Medical
               Science, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
FEATURES   Location/Qualifiers
     source   1..789
                     /organism="Cricetulus griseus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10029"
                     /cell_type="CHO"
     CDS   4..469
                     /codon_start=1
                     /product="nuclear DNA-binding protein CID"
                     /protein_id="AAP43113.1"
                     /db_xref="GI:31095452"
                     /translation="MAGGEMNEDYPVEIHESLSALESSLGAVDDMLKTMMSVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITD
                     KKKAAKLDRGAASRFVKNALWEPKQKNTPNVANKGKSKH"

ORIGIN

Best Local Similarity   54.3%;
Matches 680; Conservative 0; Mismatches 87; Indels 13; Gaps 5;

```
Qy      36 TATTTCCCGGAGACAGGGGTCCACGGGTATTGAGTGTGGTTCACAATGGCCAGGTGAAGAAATG 95
               ||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db       2 TCTTTCCCGGAGACAGGGCTCCACGGGTCCTGAGTTGGTCAAAATGGCAGGTGGAGAAATG 61

Qy      96 AATGAAGATTATCCCGTAGAAATTCACGAGTCTTTAACAGCCCTGGAGAGCTCCCTGGGT 155
               |||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db      62 AATGAAGACTATCCCGTAGAGATTCATGAGTCTTTATCAGCCCTAGAGAGCTCCCTGGGT 121

Qy     156 GCTGTGGACGACCATGCTGAGACCATGATGAGCTGTTTCTTTAGAAAACGAGTTGTTCAGGAG 215
               |||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db     122 GCTGTGGATGACATGCTGAAGACCATGATGAGTCTGTTTCTTTAGAAAATGAGTTGTTGCAGAAG 181

Qy     216 TTGGACCCATTGGAACAAGCAAAGGTGGATTTAGTTTCTTCGCATACACCTTAAATTCAATG 275
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 ...
```

```
Db     182 TTGGACCCACTGGAGCAAGCAAAAGGTGGATTTAGTTTCTGCATACACATTGAATTCAATG 241
               |||| |||| |||| ||||| |||||| ||||| |||| |||| |||| |||| ||||

Qy     276 TTTTGGTTTATTTGGCAACTCCAAGAGAGTTAATCCCAAAAGAGCATCCAGTTGAAGCAGGAA 335
               |||| ||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db     276 TTTTGGGTTTATTTGGCAACTCCAAGAGAGTTAATCCCAAAAGAGCATCCAGTTGAAGCAGGAA 335

Qy     336 CTGGAAAGAATCAGAGTCTACATGAACAGAGAGTTAAAGAAAATAACAGACAGAGAAGAAGGCT 395
               |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db     302 TTGGAAAGAATCAGAGTCTACATGAACAGAGAGTTAAAGAAAATAACAGACAGAGAAGAAGGCT 361

Qy     396 GCCAAGCTGGACAGAGGTGCTGCTTCGCAGATTTGTCAAGAAGGCACTCTGGGAACCCAAA 455
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 GCCAAGCTGGACAGAGGTGCTGCCTCGCAAGATTTGTCAAGAAATGGCGCTCTGGGAACCCAAA 421

Qy     456 CGAAAAAGCACACCAAAAGTCGGCTAATATAAAAGGGGAAAGCAAAACACTAAATCTTTTGGTTTCT 515
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 CAGAAAAACACACCAAAAGTGTTGCTAATAAGAAGTAAACATTAAATCTTCTGGTTTCT 481

Qy     516 GATGTACATGTTTTCAAAAAAGTACATCCTTTTTTTTAATCAGTT---TACAAGTGTAGTTATGT 572
               |||| |||| ||||||||| |||||| || ||||||  |||||       ||||| ||| ||||
Db     482 GATGTACATGTTTTCAAAAAAGTACATCGTTTTTTACATTTACAAAAATAGGTAGTTGTGT 541

Qy     573 GACCAGTGGTGTTGTTTAAATGAATTCCTTTGGAATTCATGTATAAAATTACACATTACAT 632
               |||| ||||||||| ||||| |||||||||| |||||||||||| ||||| |||| ||||
Db     542 GGCAATGCAAGTGTTTTTAAATGGAATTCCTTTGAAATTCATGTATAAAATTACACATTAAAT 601

Qy     633 TTG------TGATACTGAATCTTTTTTTTTGGCTGAGAAGAGATTAAGTTGTCTTTGTTGAT 685
               |||       ||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 TTGGGTGATACTGAATCTTTTTTTTTGGCTGAGAAAAGAGATTAAGTGAGAAAAGAGATTAAGTTGTCATTGATT 661

Qy     686 TTT-CATATAAAGCATCAT-GATGTGTTTAAATATTGTTAAGATAATTCTATAAGCAGTGTG 743
               |||  |||||||||||||  |||||||||||||||| ||||| ||||||||||||||||
Db     662 TTTCCATATAAAGCATCCTAGATGTGTTTAAATAGTGTTTAAGGTATTCTATAAGCAATGTG 721

Qy     744 AAAATCCAAAATGTT-CTCTGTAAACATTTGTTAGTGTTTGAAATGAACAATGATATTATGAA 802
               ||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db     722 AAATCCAAATGTTCTTTTGTAAACATTTGTTAGTACTTTAAATGAATAATGATATTATGCA 781
```

RESULT 11

LOCUS   AC123163   209355 bp   DNA   linear   HTG 13-MAY-2003
DEFINITION   Rattus norvegicus clone CH230-165P8, *** SEQUENCING IN PROGRESS
               ***, 3 unordered pieces.
ACCESSION   AC123163
VERSION   AC123163.5 GI:30579265
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM   Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE   1 (bases 1 to 209355)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
               Muzny,D.Marie., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
               Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
               Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
               Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
               Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
               Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
               Cardenas,V., Carter,K., Cavazos,J., Ceasar,H., Center,A.,
               Chacko,J., Chavez,D., Chen,J., Chen,R., Chen,Z., Chu,J.,
               Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
               Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
               Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
               Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
               Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
               Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
               Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
               Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
               Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
               Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
               Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
```

```
TITLE     Direct Submission
JOURNAL   Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 2  (bases 1 to 209355)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE 3  (bases 1 to 209355)
AUTHORS   Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
          Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
          Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
          Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
          Liu,J., Liu,W., Liu,J., London,P., Longacre,S., Lopez,J., Lozado,R.J.,
          Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
          Maheshwari,M., Mapua,P., Martin,R., Martin,M., Malloy,K., Mangum,A.,
          Mangum,B., Mahindarine,M., Mahmoud,M., McNeill,T.Z., Meenen,E.,
          Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Martinez,E.,
          Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
          Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
          Nankervis,C., Neal,J., Newton,N., Nguyen,N., Norris,S.,
          Nwaokelemeh,O., Okwuonu,G., Olaripunsagoon,A., Pal,S., Parks,K.,
          Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
          Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.L.,
          Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
          Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
          Rives,C., Rockey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
          Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
          Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
          Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
          Steimle,M., Strong,R., Sutton,G., Svatek,A., Tabor,P., Taylor,C.,
          Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
          Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
          Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
          Williams,G., Willeon,R., Wleczyk,R., Wooden,H., Worley,K.,
          Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
          Yu,F., Zhang,J., Zhou,J., Zhao,X., Zhao,S., Dunn,D., von
          Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
          Weinstock,G. and Gibbs,R.A.
TITLE     Direct Submission
JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On May 13, 2003 this sequence version replaced gi:25188596.
COMMENT   The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.

          --------- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          --------- Project Information
          Center project name: GMNO
          Center clone name: CH230-165P8
          --------- Summary Statistics
          Assembly program: Atlas 3.0;
          Consensus quality: 190762 bases at least Q40
          Consensus quality: 194365 bases at least Q30
          Consensus quality: 197548 bases at least Q20
          Estimated insert size: 204166; sum-of-contigs estimation
          Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
          ---------------------------------------------------------------
        * NOTE: Estimated insert size may differ from sequence length
        *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank-draft-data.html).
        * NOTE: This is a 'working draft' sequence. It currently_
        *   consists of 3 contigs. The true order of the pieces
        *   is not known and their order in this sequence record is
        *   arbitrary. Gaps between the contigs are represented as
        *   runs of N, but the exact sizes of the gaps are unknown.
        *   This record will be updated with the finished sequence
        *   as soon as it is available and the accession number will
        *   be preserved.
        *      1   182365: contig of 183365 bp in length
        *   182366   182465: gap of unknown length
        *   182466   208006: contig of 25541 bp in length
        *   208007   208106: gap of unknown length
        *   208107   209355: contig of 1249 bp in length.
FEATURES             Location/Qualifiers
     source          1..209355
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-165P8"
     misc_feature    1..1679
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3627..4180
                     /note="clone_boundary
                     site:EcoRI
                     clone_end:Sp6
                     end sequence:BH277959"
     misc_feature    16474..18153
                     /note="wgs_contig"
ORIGIN
```

```
Query Match          44.9%;  Score 467.2;  DB 2;  Length 203355;
Best Local Similarity  87.7%;  Pred. No. 2.7e-83;
Matches 614;  Conservative   0;  Mismatches  73;  Indels  13;  Gaps  9;

Qy    338 GGAAAGAATCAGAGTCTCACATGAACACAGAGTTAAAGAAAATAACAGAGAAGAAGAGGCTGC 397
          ||||||||||||||||||||||||||||||||| ||||||||||||||||| || |||||||||
Db 192856 GGAAAGAATCCGAGTCTACATGAACACAGAGTTAAAGAAAATAACAGACAAGAAGAGGCTGC 192915

Qy    398 CAAGCTGGACAGAGGTGCTGCTTCGAGATTTGTCAAGAAGGCCACTCTGGGAACCCAAACG 457
          ||| ||||||||||||||  ||||||||| ||||||||||||||||| |||| ||||||||
Db 192916 CAAGCTGGACAGAGGGGCGCCGCGTCGACATTGTCAAGAAGGCCACTCTGGGAACCCAAACC 192975

Qy    458 AAAAAGCACACCAAAAGTGGGCTAATAAAGGGAAAAAGCAAACAC---TAATCTTTTGGTTT 514
          |||||||||||||||||| || |||||||  ||||||||||||||    |||||||||||||
Db 192976 AAAAAGCACACCAAAAGTGGCTAATAAAAGGGAAAAAGTAAACACTAATAATCTTTTGGTTT 193035

Qy    515 TGATGTACATGTTTCAAAAAGTACATCCTTTTAATCAGTTTACAAATGTGTATGTGTGA 574
          ||||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||
Db 193036 TGATGTACGTGTTTCAAAAAGTACATCCTTTTA-ATCAGTTTACAGATGTACTTATGTGA 193094

Qy    575 CCATGTGGGTGTTTAAATGGATTCCTTTTTGAAATTCATGTTGTAAAATTTACACATTACA-TT 633
          |||||||||| ||||||||||||||||||||||||||||||||| |||||||||||| || ||
Db 193095 CCATGTGGGTGTTTAAATGGATTCCTTTTTGAAATTCATGTATAAAAATTTACACATTACA-TT 193154

Qy    634 TGTGATACTGAATC---TTTTTTTGCTGAGAAAGATTAAGTTGTGTTCTTTGTTTGATTTTCAT 691
          ||||||||||||||||  |||| || || || ||||||||||||||||||||||||||||||||
Db 193155 TGTGATTACTGAATCTTTTTTTTTTTTTGCTGAGAAAGATTAAGTTGTCTTCGTTTGATTTTTCCT 193214

Qy    692 ATAAAGCATCATGATGTGTTTAATATTGTAAGATAATTCTATAAGCAGTTGTGAAATCCAA 751
          ||| |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db 193215 ATAAAGCATCATGATGT-TTTAATTGAAGAGATATCTATAAGCAACTGTGAAAATCCAA 193273

Qy    752 ATGTTCTCTGTAAACAIATTTGTTAGTGTTTGAAATGAACAAATGAATATATGACGTGTGAT 811
          ||||||||||| |||||| |||||||||||||||||||||||| ||||||||||||||||||
Db 193274 ATGTTCTTTGTAAACGTTTGTAGTGTTTTAAATG--TAGTGTTATGATGAAGTGTCCTGT 193331

Qy    812 CTGTAGACCTCGAGGTGTAAGGACACATTCGTTTCACGAATGATGAGAAAATGTAATGACTT 871
          |||||||| ||||  ||||||| ||||||||| |||| | |||||||||||| ||||||||
Db 193332 CTGTAGACTGGAGTTGTTGTTGAGAGGACAATCGTTTCACAATGATGAGAAAAATGTAATGACTT 193391
```

```
Qy    872 AAAATACCCACTCT-GTTTCTCGTTCAGTTCAGTTGTCAACATAGTTTCGGGATTTTTTTTTTTT 930
          |||||||||||||| |||||||||||||||||||| ||||||| |||| |||||||| |||||||
Db 193392 AAAATACCCACTCTGGTTTCTCGTTCAGTTCAGTTCAACATAGTTTC-TGTGATTTTTTTTTTA 193450

Qy    931 TTGAGTAATTCGTCGTGATATTCAAAGT-CAAAATTGAAACCTTAAGGCTGTACTTTAA 989
          ||||||||| || ||||||| |||||||| |||||||||||| |||||||||| |||||
Db 193451 GGAGTAATTGCTGTCGTGATGTTCACAGTACAAAATTGAAACCTTCAGGCTGTATTTAA 193510

Qy    990 TTCTTCATGTTCCATTTAAAATAAAAAATAGTTCTCATTAACT 1029
          ||||||||||||||||||||||||||||| |||||||||| |
Db 193511 TTCTTCATGTTCCATTTAAAATAAAAAATAATTCTCATTCATT 193550
```

RESULT 12

```
ACCESSION   AC097131
LOCUS       AC097131             221055 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-49P11, ***, 4 unordered pieces.
ACCESSION   AC097131
VERSION     AC097131.7  GI:30520752
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 221055)
  AUTHORS   Muzny,D.Marie, Metzker,M.Lee, Abramzon,S., Adams,C., Alder,J.,
            Allen,H., Allen,H., Alsbrooks,I., Amin,P., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,N., Hamil,C., Hamilton,K., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,B., Hsiyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindadtne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Mlioaavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsaagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
            Puazo,J., Quiroz,J., Rachlin,E., Reeves,K., Regier,M., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
```

```
            Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G., and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221055)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 221055)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:25091251.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            --------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            --------------- Project Information
                Center project name: GHXP
                Center clone name: CH230-49P11
            --------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 19751 bases at least Q40
                Consensus quality: 200098 bases at least Q30
                Consensus quality: 202942 bases at least Q20
                Estimated insert base: 205057; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            ------------------------------------------------------------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE:This is a 'working draft' sequence. It currently
            *   consists of 4 contigs. The true order of the pieces
            *   is not known and their order in this sequence record is
            *   arbitrary. Gaps between the contigs are represented as
            *   runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            *   as soon as it is available and the accession number will
            *   be preserved.
            *        1        42645: contig of 42645 bp in length
            *    42646        42745: gap of unknown length
            *    42746       173337: contig of 130592 bp in length
            *   173338       173437: gap of unknown length
            *   173438       218852: contig of 45415 bp in length
            *   218853       218952: gap of unknown length
            *   218953       221055: contig of 2103 bp in length
FEATURES             Location/Qualifiers
     source          1..221055
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-49P11"
     misc_feature    3702..4811
                     /note="wgs_contig"
     misc_feature    51725..52966
                     /note="wgs_contig"
```

```
ORIGIN

         Query Match          44.9%;   Score 467.2;  DB 2;  Length 221055;
         Best Local Similarity 87.7%;   Pred. No. 2.7e-83;
         Matches 614; Conservative  0; Mismatches 73; Indels 13; Gaps  9;

Qy   338 GGAAAGAATCAGAGTCTACAGAGACACAGAGCTTAAGAAAATAAACAGACAAGAAGAAGGCTGC 397
         |||||||||||||||||||||||||||||||||| ||| |||||||||||||||||||||||||
Db 158779 GGAAAGAATCCGAGTCTACATGACACAGAGTTAAAGAAAATAAACAGACAAGAAGAAGGCTGC 158838

Qy   398 CAAGCTGGACAGAGGTGCTGCTTCGGAGATTTGTTCAAGGAGGCACTCTCGGGGACCCAAACG 457
         ||||||||||||| ||| |||||||||||||||||||||||||||||||||||||||||||||
Db 158839 CAAGCTGGACAGAGAGGGCCGCGCTCGAGATTTGTTCAAAAATGCACTCTCGGGACCCAAACC 158898

Qy   458 AAAAAAGCACACCAAAAAGTGGGCTAATAAAAGGGAAAAGGCAAACAC---TAATCTTTTGGTT 514
         ||||| |||||||||||| || | |||||| | ||||| ||||||||||   |||| ||||||||
Db 158899 AAAAAAACACACCAAAAAGTGGCTAATAAAGTAAAACACTAATAATCTTTTGGTTT 158958

Qy   515 TGAGTACATGTTTTCAAAAAGTACACATCCTTTTTAATCAGTTTACAATGTAGTTATATCTGA 574
         ||||||||||||||||||||||||||||||| ||||||||||||||| |||||| |||||
Db 158959 TGAGTAGGTGTTTTCAAAAAGTACATCATTTTT-ATCAGTTTACACAGTGTACTTATGTGA 159017

Qy   575 CCATGGTGGTGTTTAAATGGATTCCTTTTGTTTCAGTATAATTTACACATTACA-TT 633
         ||| ||| ||||| ||| |||| |||||||||||||||| ||| ||| |||||| ||
Db 159018 CCAATGTGGCTGTTTAAATGGATTCCTTTTGTTTCAGTATAAATTTACACATTAAATTT 159077

Qy   634 TGTGATACTGAATC--TTTTTTTGTCGAGAAAGAATTAAGTTGTCTCTTTGTTCATTTTCAT 691
         |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db 159078 TGTGATACTGAATCTTTTTTTTTTGTCTAGAAAGAATTAAGTTGTCGTTATTGATTTTCCT 159137

Qy   692 ATAAAGCATCAGATGTGTGTTTAATATGTAAGATATTCTATAAAGCAATTGTTGTGAAATCCAA 751
         ||||||||||| || ||||||| |||||||| ||||||||||||| |||||||||||| |||||
Db 159138 ATAAAGCATCATGATGT-TTTAATATTGAAAGATATTCTATAAAGCAATTGTTGTGAAATCCAA 159196

Qy   752 ATGTTCTCTGTAAACATTTGTAGTGTTGAAATGAACAATGAGTATTATGAGAGTGTGCTAT 811
         ||||||||  ||||||||||| ||| |||||||||||||||||||||||||| ||| |||||
Db 159197 AATGTTCTTTGTAAAACGTTTGTAGTGTTTAAATG--TAGTGTATGAGTGAAAGTGTGCTGT 159254

Qy   812 CTGTAGACCTCGAGGTGTAAGGACACATTGTTTTTTCAGTATAATGATGAGAAATACAGTGACTT 871
         |||||| ||||||| ||||||||||||||||||||||||||| || ||||||| |||||| |||
Db 159255 CTGTAGACTGGAGTGTTAGAGGACAATCTGTTTTCACAATGAACATAAGACTT 159314

Qy   872 AAATACCCACTCT--GTTTCCTGTTCAGTTAGTTCAACATTGTTTC-TGTGAATTTTTTTTTTTT 930
         ||||||||| || |||||||||||||||| ||||||| |||||||   |||||||||||||||
Db 159315 AAATACCCACTCTGGTTTCTGTTCAGTACAACATGTTTC-TGTGAATTTTTTTTTTTA 159373

Qy   931 TTGAGTAAATTCTGTCTCTTGATATTCAAAGT-CAAAATTGAAACCTTAAGGCTGTACTTTAA 989
         |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db 159374 GGAGTAATTGCTGTCTTGATCTTCACAGTACAAAAATTGAAACCTTCAGCGTGTTATTTAA 159433

Qy   990 TTCTTCAGTTCCATTTAAAATAAAATATATTCTCATTCATT 1029
         |||||||||||||||||||||||||||||||||||||||||
Db 159434 TTCTTCATGTTCCATTTAAAATAAAATATATTCTCATTCATT 159473
```

```
FEATURES
     source         1. 1156
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
```

```
ORIGIN

         Query Match          41.7%;   Score 433.8;  DB 6;  Length 1156;
         Best Local Similarity 71.1%;   Pred. No. 2.2e-76;
         Matches 662; Conservative  0; Mismatches 252; Indels 17; Gaps  6;

Qy    66 GAGTTGGGTCACAATGGCAGGTGAAGGAAATGAATGAAGATTATCCCGTAGAAAATTCACGAG 125
         |||||| || ||||| ||||||| |||||||||||||||| ||||||||||||||||||||||
Db   106 GAGCTGGCCATAATGGCAGGTGAAGGAAATTAATGAAGGACTATCCAGTAGAAAATTCACGAG 165

Qy   126 TCTTTAACAGGCCCTGGAGAGCTCCCTGGGTGCTGTTGGGACGACATTGCTGAAGACCATGATG 185
         |||| ||||||||||||||||||| ||||| |||||||||||| |||||||||||||||||||
Db   166 TATTTGTCAGGGTTTGAGAATTCCAGTGGTGTCGTGTTGGGACGAGTAATCAGAAGACCATGATG 225

Qy   186 GCTGTTTCTAGAAAACCAGTTGTTGTTGCAGAGTTGGAAGTTGGAACAACAAGCAAAGGTGGAT 245
         ||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||| |||
Db   226 TCTGTTTCTAGAAATGAGTTGTTGCAGAAGTTTGGAATCCACTTGAACAACAACCAAAAGTGGAT 285

Qy   246 TTAGTTTCTGCATACACCTTAAAATTCAAATGTTTGGGTTTTATTGGCAACTCAAGGAGTT 305
         ||||||||||||||| ||||||||||||||||| ||||||||||||| |||||||||||||
Db   286 TTGGTTTCTCGATACACATTCAAAATTCAAGTTTATTGGCAACCAAGGAGTT 345

Qy   306 AATCCCAAAGACATCCAGTGAAGCGGAGCTGGAAGAAAATCAGAGTTCACAGTGAACAGA 365
         |||||||||||||||||||||| |||||||||||||||||||||||||||||||| ||||
Db   346 AATCCTAAGGAACATCCAGTAAAACAGGAAAATTGGAAAGAAATCAGAGTGTATATGAAACAGA 405

Qy   366 GTTAAAGAAATAACAGACAAGAACAGAGAGAAGGGCTGCCAAGCTGGATGGTGCTCGCTTCGAGA 425
         ||||||||||||||||||||||| ||||||||||||| ||||||||||||||||||||||||
Db   406 GTCAAGGAAATAACGACAAGAAAAAAAGGCTGGCAAAAGGTTGCAGCTTCAAGA 465

Qy   426 TTTGTCAACAAGAGGCACTCTGGGACCCAAACACAAAAAGTAGTGCTAATAA 485
         |||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   466 TTTGTAAAAAAATGCCCTCTGGGAACCAAAATCGAAAAAGTTGCCAATAAA 525

Qy   486 GGGAAAAAGCAAACACTAATCTTTTGGTTTTTGATGTAGTACATGTTTCAAAAAAGTACA-TCCT 544
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||||
Db   526 GGAAAAAGTAAAAAGTTAACTTTTTGGTTTTTGATGTAGTACACATATTCAAAAAAGTACATAAT 585

Qy   545 TTTTAAATCAGTTTACAAGTAGTTATGTGACCAATGGGTGTTTAAATGGATTCCTTTGG 604
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   586 ATGTAAATCACAGTAATAATAAAGTAAATACTTCCTCTCCAAAGAATCATTATCTTTAT 645

Qy   605 AATTCAATGTATAAAATTTACACATTACACATTTGTGATACTGAATCTTTTTTTTGCTGAGAAA 664
         |||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   646 GAT-----TAGCACTGAGGATTTTAACATTTGTGATATATATATTTATATTTAAATTACCAT 700

Qy   665 GATTAAGTTGTCTTTGTTGATTTTCATATAAAGCATAAGATGTGTTTATGGTTTTATATTGTAAGA 724
         ||||||||||||||||||| |||||||||||||||||||||
Db   701 CTCTTGATGAGACTCTTTATTCTTTTATAGGTCAGTCCTGCAAGGTGTTTATTTTATAAGC 760

Qy   725 TATTCTATAAAGCAGTTGTGAAATCCAAATGTTCTCTGTAAACATTTGTAGTGTTTGAAAT 784
         ||||||||||||||||||||||||||||
Db   761 AGCTGTGAAA--------TTTAAGTGAAATGTTCTTGTAAACATTTGTACTATTTTTAAAT 813

Qy   785 GAACAATGATTATGAAGTGTTGCTATCTGTAGACCTCGAGGTGTAAGGACACATTGTTTT 844
         ||| ||||||| ||||||||||||||||||||| |||||
Db   814 GAATAATGACCTTATGAAGTAATGCTATCTCGTTAG-GCTGAAATTATAGGTACATCCTTGTTTT 872
```

```
Qy  845 CAGTAATGATGAGAAATACAGTGACTTAAATACCACTCTGTTCTCGTGTTCAGTTAGTTCA 904
        ||||     | |||| ||||| || |||||||||||     || |||      | |||| |
Db  873 CACTA--TATGATATTAAGAAAGCGTGAAATGACTTAAATGTTCATTTTTCTGTATAGA    930

Qy  905 ACATGTTTCCGTGATTTTTTTTTTTTTGAGTAATGACTTAAATGTTCAAAGT-CAAA     963
        || ||| || ||| ||||| || |||| ||| || || |||| ||||||||| ||||
Db  931 TACTTTATCATGTTTTTCATGATTTTTAGGAATTACTGCTTGTTGTGATAATTCAAAGTGTGAA 990

Qy  964 ATTGAAAACCTTAAGGCTGTACTTTAATTCTT 994
        |||| |  |  | ||| ||||||||||||||||
Db  991 ACTAAAAAGTTTATGGTTGTACTTTAATTCTT 1021
```

RESULT 14
AX009361

LOCUS   AX009361   1156 bp   DNA   linear   PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO9963071.
ACCESSION   AX009361
VERSION   AX009361.1 GI:9996662
KEYWORDS
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Rothbarth,K., Werner,D. and Stammer,H.
  TITLE   Method for triggering apoptosis in cells
  JOURNAL   Patent: WO 9963071-A1 09-DEC-1999;
            DEUTSCHES KREBSFORSCH (DE); ROTHBARTH KARSTEN (DE); WERNER DIETER
            (DE); STAMMER HERMANN (DE)
FEATURES   Location/Qualifiers
     source   1..1156
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS   118..543
                     /codon_start=1
                     /protein_id="CAC07450.1"
                     /db_xref="GI:9996663"
                     /note="unnamed protein product"
                     /translation="MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELL
                     QKLDPLEQAKVDLVSAYTLNSMFVVYLATQGVVNPKEHVKQELERIRVYMNRVKEITD
                     KKKAGKLDRGAASRFVKNALWEPKSKNASKVANKGKSKS"
ORIGIN

Query Match   41.7%; Score 433.8; DB 6; Length 1156;
Best Local Similarity   71.1%;
Matches 662; Conservative   0; Mismatches 252; Indels 17; Gaps   6;

```
Qy   66 GAGTTGGTCACAATGGCAGGTGAGAAATGAATGAAGATATCCCGTAGAAATCACGAG 125
        |||||||||||||||| |||||| ||||| || || || |||| |||| |||||||
Db  106 GAGCTGGCCATAATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAG 165

Qy  126 TCTTTAACAGCCCTGGAGAGCTCCTGGGGTGCGTGTGGACCACATGCTGAAGACCATGATG 185
        ||||  |||| |||| || |||  || | ||| ||||| || |||| |||||| ||||||
Db  166 TATTTGTCAGCGTTTGAGAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATG 225

Qy  186 GCTGTTTCTTAGAAACGAGTTGTTGGCAGAAGTTGGACCCATTGGACAAGCAAAGGTGGAT 245
        ||| || | | || |  |||||||| ||| || |||  |  || ||| | |||| ||||
Db  226 TCTGTTTCTAGAAATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGAT 285

Qy  246 TTAGTTTCTGCATACACACCTTAAATTCAAATGTTTATTGGCAACTCAAGGAGTT 305
        ||| ||||| ||| |||| ||   |||||| ||||| ||||||   || ||||||
Db  286 TTGGTTTCTGCATACACATTAAATTCAAATGTTTATTTGGCAACTCAAGGAGTT 345

Qy  306 AATCCAAAGAGCATCCACTGAGAAAGAATCACGAGTTCTACATGAACAGA 365
        || |||| |  ||| || ||| | |||| |||||| || |||| |||||
Db  346 AATCCTAAGGAACATCCAGTAGAAAGCACAGGAAGTCAGTAATCAGAACAGA 405

Qy  366 GTTAAAGAAAATAACAGCAAGAAGGAGAAGAGGCTGCCAAGCTGGAAGCTTTGCTTCGAGA 425
        ||| ||||||||||||| | ||| ||| ||||||| | |||||   | | |||| |||||
Db  406 GTCAAGGAAAATAACAGACAAGAAGAAAAGGGTTGCAGCGTTCAAGA 465

Qy  426 TTTGTCAAGAGGACACTCTGGGAGAACCCAAACGAAAAAAGCACACCAAAAAGTGGCTAATAA 485
        |||| |||| |  ||| ||||| || || ||| || |||||| || | ||||||| |||||
Db  466 TTTGTAAAAAATGCCCTCTGGGAACCAAAATGTAAAAATGTGCCAATAA 525

Qy  486 GGGAAAAGCAAACACTAAATCTTTGGTTTGTGATGTAGACAGTTTTCAAAAAGTACA-TCCT 544
        || ||||||| || || ||| || || | || || ||| ||||||||| ||||||| |||
Db  526 GGAAAAAGTAAAAGTTAACTTTTTGGTTTTGATGTAGACAGTTTTCAAAAAGTACA-TCCT 585

Qy  545 TTTAATCAGTTACAAATGTAGTTAGTGACCATGTGGTGTTTAAATGAATTCCTTTGG 604
        ||||||||| ||||||||| || |||| |||| ||||| ||||| |||||| | ||
Db  586 ATGTAATCACAGTAGAATAGTGAAAAACTTCCTCTCCAAAGATCATTATCTTTATT 645

Qy  605 AATTCATGTATAAAATTTACACATTACATTACTTTGTTGGTACTGAATCTCTTTTTTGCTGAGAA 664
        |||||| || ||||| ||||| |||||||| |  |||| |||| || |   | ||||| || |||
Db  646 GAT-------TAGCACTGAGGATTTTAACATTGTTGATTATATATATTATTTATTCACCAT 700

Qy  665 GATTAAGTTGTCTTGTGTTGATTTCATATAAAGCATCATGATGTGGTTTTCAAAATGTGAGA 724
        ||||||||| || ||||| || |||| |||||| || |||| | | ||||||| |||||
Db  701 CTCTTGATGAGACTCTTATTTTCTTATATAGGTCAGTGTCTTGGAAGTACCATTTTATAGC 760

Qy  725 TATTCTATAAGCAGTTGTGAAATCCAAAATGTTCTCTGAAATGTTGTCTGTGAAAT 784
        |||||||||| || ||||||||| |||||| ||||| |||||| ||||||| ||||
Db  761 AGCTGTGAAA-------TTTAAGTGAAAATGTTCTTTGTGTAAACATTTTGTGTACTATTTTTAAAT 813

Qy  785 GAACAATGATATTATGAAAGTGTGCTATCTGTAGACCTCGAGGTGTAAGGACATTTGTTTT 844
        |||  ||| ||||||| |||||| || ||||| |||  |||| |||| |||  |  ||||
Db  814 GAATAATGACCTTATGAAAGTAGCTATCTGTAG-GCTGAAATTATAGGTACATCTGTTTTT 872

Qy  845 CAGTAATGATGAGAAATACAGTGACTTAAATACCACTCTGTTCTCGTGTTCAGTTAGTTCA 904
        ||||     | |||| ||||| || |||||||||||     || |||      | |||| |
Db  873 CACTA--TATGATATTAAGAAAGCGTGAAATGACTTAAATGTTCATTTTTCTGTATAGA    930

Qy  905 ACATGTTTCCGTGATTTTTTTTTTTTTGAGTAATGACTTAAATGTTCAAAGT-CAAA     963
        || ||| || ||| ||||| || |||| ||| || || |||| ||||||||| ||||
Db  931 TACTTTATCATGTTTTTCATGATTTTTAGGAATTACTGCTTGTTGTGATAATTCAAAGTGTGAA 990

Qy  964 ATTGAAAACCTTAAGGCTGTACTTTAATTCTT 994
        |||| |  |  | ||| ||||||||||||||||
Db  991 ACTAAAAAGTTTATGGTTGTACTTTAATTCTT 1021
```

RESULT 15
AX329809

LOCUS   AX329809   1172 bp   DNA   linear   PAT 09-JAN-2002
DEFINITION Sequence 318 from Patent WO0194629.
ACCESSION   AX329809
VERSION   AX329809.1 GI:18102787
KEYWORDS
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.B., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE   Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 318 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES   Location/Qualifiers
     source   1..1172
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match   41.4%; Score 431; DB 6; Length 1172;
Best Local Similarity   70.9%;
Matches 660; Conservative   0; Mismatches 255; Indels 16; Gaps   6;

```
Qy   66 GAGTTGGTCACAATGGCAGGTGAGAAATGAATGAAGATATCCCGTAGAAATTCACGAG 125
        |||||||||||||||| ||||||  ||||| || || || |||| |||| |||||||
Db  106 GAGCTGGCCATAATGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAG 165
```

```
Qy   126 TCTTTAACACCCTGGAGAGCTCCCTGGGTGCTGTGTGGACGACAATGCTGAAGACCAATGATG 185
             |||                                                              
Db   166 TATTTGTCAGCGGTTTGAGAATTCCATTGGTGCCTGTGGATGAGATGCTGAAGACCAATGATG 225

Qy   186 GCTCGTTTTCTTAGGAAACGAGTTGTTGCAGAAGTTGGACCCATTGGAACAAAGCAAAAGTGGGAT 245
             ||||||||                                                           
Db   226 TCTGTTTTCTTAGAAATGAAGTTGTTGCAGAAGTTGGATCCACTTGGACAAGCAAAAAGTGGAT 285

Qy   246 TTAGTTTCTGCATACACCTTAAATTCAAATGTATGTTTCGGGTTTTATTTGGCACTTCAAGGGAGTT 305
             ||     ||                                                           
Db   286 TTGGTTTCTCCATACACATTAAATTCAAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTT 345

Qy   306 AATCCCAAAGAGCAATCCAGTGAAGCAGGGAACTGGAAAGAATCAGAGTCTCACATGAACAGA 365
             |||                                                             
Db   346 AATCCTAAGGAACAATCCAGTAAAACACAGGAATTGGAAAGAATCAGAGTATTATGAAACAGA 405

Qy   366 GTTAAAGAAATAACGACAAGAGAAGGGCTCCCAAGCTGACACGAGGGTGCTGCTTCGAGA 425
             ||||||                                                       
Db   406 GTCAAGGAAATAACAGAAAGGCTGGCAGAGAGGTGCAGCTTCAAAGA 465

Qy   426 TTTGTCAAGAAGGCACTCTCTGGGAACCCAAAGGAAAAAGCACACCAAAAAGTGGGCTAATAAA 485
             ||||||                                                             
Db   466 TTTGTAAAAAATGCCCTCTGGGAACCAAAAATCGAAAAAATCGAAAAAGTTGCCAATAAA 525

Qy   486 GGGAAAAGCAAAACACTAAATCTTTTGGGTTTTGAGTACATGTTTTCAAAAAGTACA-TCCT 544
             |||||                                                          
Db   526 GGAAAAAGTAAAAGTTAACTTTTTTGGTTTTTGATGTACACATATTCAAAAAGTACACATTAAT 585

Qy   545 TTTTAATCAGTTTACAAATGTAGTTATGTGACCATGTGGTGTTTAAATGGATTCCTTTTGG 604
             |||||                                                         
Db   586 ATGTAATCACAGTAATATGTAAAGCTAAATACTTCCTCTCCAAAAGATCATTATCTTTATT 645

Qy   605 AATTCATGTATAAAATTTACACATTACATTTCATGTTGTGATATCTTTTTTTTGCTGAGAAA 664
             ||                                                              
Db   646 GAT-----TAGCACTGAGGATTTTTACACTTGTGATATATTATATAATTTATTACCAT 700

Qy   665 GATTAAGTTGTCTTGTTGATTTTCATATAAAGCATCAATGATGTGTTTAATATTGTAAGA 724
             ||                                                           
Db   701 CTCTTGAATGAGACTCTTATTTCTTTATAGTTAGGTCAGTCTTGCAAGTACCATTTTATTAAGC 760

Qy   725 TATTCTATAAGGCAGTTGTGAAATCCAAAGTTGTTCTCTGTAAACATTTGTAGTGTTTGAAAT 784
             ||                                                              
Db   761 AGCTTGTGAAA-------TTTAAGTGAAAGTTCTTTTGTGTAAACATTTGTACTATTTTAAAT 813

Qy   785 GAACAATGATAATTATGAAGTGTGCTATCTGTAGACCTTCGAGGTGTAAGGACACATTTTGTTTTT 844
             ||                                                                 
Db   814 GAATAATGACCTTATGAAGTATGCTACCTGTAG-GCTGAAATTATAGGTACAACTCGTTTTT 872

Qy   845 CAGTAATGAATGAGAAATACAGTGGACTTAAATACCCACTCGTTTCTCTTCAGTTAGTTCA 904
             ||                                                            
Db   873 CACTAATGATAATTAAGGATTAAGCAAAAGGG-TGAAATGACTTAAATGTTCATTTTTTCTGTAGTAGA 931

Qy   905 ACATGTTTCGTGATTTTTTTTTTTTTTTTTGAGTAAATCTGTCTTGATATATCAAAAGT-CAAA 963
             ||                                                               
Db   932 TACTTTATCAATGTTTTCATGATTTTTAGGAAATTACTGCTTTTGTTGTGATTATTCAAAAGTGTGAA 991

Qy   964 ATTGAAACCTTAAGGGCTGTACTTTAATTTCTT 994
             ||                                
Db   992 ACTAAAAGTTTATGGTTGTACTTTAATTCTT 1022
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: January 4, 2005, 03:11:18

Title: US-09-701-618A-3
Perfect score: 1040
Sequence: 1 cagaagccgtcatggcgcgt..........tcattaacctctgatggaaaa 1040
Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_23Sep04:
            1: geneseqn1980s:*
            2: geneseqn1990s:*
            3: geneseqn2000s:*
            4: geneseqn2001as:*
            5: geneseqn2001bs:*
            6: geneseqn2002as:*
            7: geneseqn2002bs:*
            8: geneseqn2003as:*
            9: geneseqn2003bs:*
           10: geneseqn2003cs:*
           11: geneseqn2003ds:*
           12: geneseqn2004s:*

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1038	99.8	1038	3	AAZ43928	Murine C1
2	431	41.4	1172	6	ABL61981	Colon ade
3	431	41.4	1172	6	ABK84717	Human CDN
4	431	41.4	1172	10	ADF81400	Leukaemia
5	431	41.4	1172	10	ADF81399	Leukaemia
6	430	41.3	1158	3	AAZ43927	Human C1D
7	418	40.2	426	12	ADJ92841	Human co-
8	392.6	37.8	586	12	ACH68585	Human gen
9	326.8	31.4	426	6	AAL43974	Mammalian
10	326.8	31.4	426	6	AAL43973	Mammalian
11	325.6	31.3	420	6	ADI28090	Human nuc
12	325.6	31.3	420	6	ADI28091	Human nuc
13	314.6	30.3	420	6	AAL43975	Mammalian
14	314.6	30.3	420	6	AAL43976	Mammalian
15	263.2	25.3	380	3	AAC01237	Human sec
16	191.2	18.4	343	10	ADB55499	Toxicity-
17	191.2	18.4	343	10	ADB50044	Primary r
18	152.8	14.7	204	12	ACH82285	Human gen
19	83	8.0	443	10	ADF79453	Leukaemia
20	66	6.3	6120	10	ABL32493	Human imm
21	62.8	6.0	17131	6	ABL33052	Human imm

22	61.4	5.9	17183	6	ABL32486	Human imm
23	60.2	5.8	65	6	ABN53492	Mouse sp1
24	59.8	5.8	8056	8	ABZ10246	Haematopo
25	59	5.7	5888	6	ABL34457	Human met
26	58.8	5.7	7544	4	AAS45300	DNA trans
27	58.8	5.7	7544	4	AAS45300	Chemicall
28	58.6	5.6	15548	6	ABK28139	Human imm
29	58.6	5.6	15548	6	ABL34155	Human imm
30	58.2	5.6	8056	8	ABZ10246	Haematopo
31	57.8	5.6	12237	6	ABL34358	Human imm
32	57.2	5.6	9021	4	AAS46325	Tumour su
33	57.2	5.5	778	6	ABQ15588	Oligonuc1
34	57.2	5.5	778	6	ABQ15589	Oligonuc1
35	57	5.5	17918	6	AAS61419	Human gen
36	57	5.5	6106	4	AAS46429	Tumour su
37	57	5.5	6106	4	AAK40031	Human che
38	57	5.5	6106	4	AAK40031	Human che
39	57	5.5	34688	8	ABQ67059	Human ang
40	56.8	5.5	8056	8	ABZ10100	Haematopo
41	56.8	5.5	8101	6	ABL33978	Human imm
42	56.6	5.4	38342	4	AAS46746	Tumour su
43	56.6	5.4	38342	6	AAS46746	Tumour su
44	56.6	5.4	50000	6	ABL56201	Signal tr
45	56.4	5.4	113515	6	ABL34174	Human imm

ALIGNMENTS

RESULT 1
AAZ43928
ID   AAZ43928 standard; cDNA; 1038 BP.
XX
AC   AAZ43928;
XX
DT   17-MAR-2000 (first entry)
XX
DE   Murine C1D cDNA.
XX
KW   C1D; murine; apoptosis; tumour; gene therapy; treatment; ss.
OS   Mus sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   76..501
FT   /*tag= a
FT   /product= "C1D"
XX
PN   DE19824811-A1.
XX
PD   09-DEC-1999.
XX
PF   03-JUN-1998; 98DE-01024811.
XX
PR   03-JUN-1998; 98DE-01024811.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Rothbarth K, Stammer H, Werner D;
XX
DR   WPI; 2000-063506/06.
DR   P-PSDB; AAY51025.
XX
PT   Inducing apoptosis by overexpressing the C1D gene, particularly for
PT   treating tumors.
XX
PS   Claim 3; Fig 2; 10pp; German.
XX
CC   This invention describes a novel method for inducing apoptosis which
CC   comprises overexpressing the C1D gene (I). The method is particularly
CC   used to treat tumors and can also be used in gene therapy. The method has
CC   no side effects on normal cells (contrast known methods of inducing
CC   apoptosis such as cytotoxins and radiation), and may be effective on

```
CC   cells resistant to conventional treatments. Overexpression of (I) is
CC   sufficient itself to induce apoptosis but the effect may be increased
CC   when used in combination with other anti-tumor methods. When cells
CC   transfected with (II) undergo apoptosis, they release factors that kill
CC   neighboring, non-transfected cells (bystander effect). This sequence
CC   encodes the murine CID protein described in the method of the invention
XX
SQ   Sequence 1038 BP; 323 A; 163 C; 219 G; 333 T; 0 U; 0 Other;

Query Match          99.8%;  Score 1038;  DB 3;  Length 1038;
Best Local Similarity 100.0%;  Pred. No. 3.5e-230;
Matches 1038;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      3 GAAGCCGTGTCATGGCGTCATCATCGTGCCACCTATTCCCGGAGACAGGCGTCCACCGT    62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAGCCGTGTCATGGCGTCATCATCGTGCCACCTATTCCCGGAGACAGGCGTCCACCGT    60

Qy     63 ATTGAGTTGGTCACAATGGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAATTCAC   122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTGAGTTGGTCACAATGGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAATTCAC   120

Qy    123 GAGTCTTTAACAGCCCCTGGAGAGTCCCTGGGTGTCGACCACATCGTGAAGACCATG      182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAGTCTTTAACAGCCCCTGGAGAGTCCCTGGGTGTCGACCACATCGTGAAGACCATG      180

Qy    183 ATGGCTGTTTCTAGAAAACAGAGTTGTTGGCAGACCATTGGAACAAGCAAAAGGTG        242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATGGCTGTTTCTAGAAAACAGAGTTGTTGGCAGACCATTGGAACAAGCAAAAGGTG        240

Qy    243 GATTTAGTTTTCTGCATACACCTTAAATTCAAATGTTTTGGGTTTATTTGGCAACTCAAGGA  302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GATTTAGTTTTCTGCATACACCTTAAATTCAAATGTTTTGGGTTTATTTGGCAACTCAAGGA  300

Qy    303 GTTAATCCCAAAAGAGCATCCAGTGAACGAGCAGGAACTCAGAGTTCTACATGAAC        362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTTAATCCCAAAAGAGCATCCAGTGAACGAGCAGGAACTCAGAGTTCTACATGAAC        360

Qy    363 AGAGTTAAAGAAATAACAGCAAGAAGAGGCTGCCAAGCTGACAGAGGTGCTCGCTTCG      422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGAGTTAAAGAAATAACAGCAAGAAGAGGCTGCCAAGCTGACAGAGGTGCTCGCTTCG      420

Qy    423 AGATTTCGTCAAGAAGGCACTCTGGGAACCCAAACACACACCAAAAGTGGCTAAT         482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGATTTCGTCAAGAAGGCACTCTGGGAACCCAAACACACACCAAAAGTGGCTAAT         480

Qy    483 AAAGGGAAAAGCAAACACTAATCTTTTGGTTTGATGACCATGTGGTGTTAAAAGGATTCCATC 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AAAGGGAAAAGCAAACACTAATCTTTTGGTTTGATGACCATGTGGTGTTAAAAGGATTCCATC 540

Qy    543 CTTTTTAATCAGTTACAATGTAGTGTGACCATGTGGTGTTTAAAATGGATTCCTTTT        602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTTTTTAATCAGTTACAATGTAGTGTGACCATGTGGTGTTTAAAATGGATTCCTTTT        600

Qy    603 GGAATTCATGTATAAATTTACACATTACAATTGTGTATCGAATCTGTTCTTTTTTTGCTGAGA 662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGAATTCATGTATAAATTTACACATTACAATTGTGTATCGAATCTGTTCTTTTTTTGCTGAGA 660

Qy    663 AAGATTAAGTTGTCTTGTTGATTTTCATATAAAGCATCATCGAATGTGTTTAATATTGTAA    722
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AAGATTAAGTTGTCTTGTTGATTTTCATATAAAGCATCATCGAATGTGTTTAATATTGTAA    720

Qy    723 GATAATCTATAAGCAGTTGTGAAATCCAAATGTTCTCTGTAAACATTTGTAGTGTTTGGAA   782
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GATAATCTATAAGCAGTTGTGAAATCCAAATGTTCTCTGTAAACATTTGTAGTGTTTGGAA   780

Qy    783 ATGAACAATGATAATATGAAGTGTGCTATCTGTAGACCTCGAGGTGTAAGACATTTGTT     842
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ATGAACAATGATAATATGAAGTGTGCTATCTGTAGACCTCGAGGTGTAAGACATTTGTT     840

Qy    843 TTCAGTAATGATGAGAAATACAGTGACTTAAATACCCACTCTGTTTCTCGTTCAGTTAGTT   902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTCAGTAATGATGAGAAATACAGTGACTTAAATACCCACTCTGTTTCTCGTTCAGTTAGTT   900

Qy    903 CAACATGTTTCGTGATTTTTTTTTTTTTTGAGTAAGTAATTCGTTCTTGATTATTCAAAGTCAA   962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAACATGTTTCGTGATTTTTTTTTTTTTTGAGTAAGTAATTCGTTCTTGATTATTCAAAGTCAA   960

Qy    963 AATTGAAACCTTAAGGCTGTACTTTAATTCTTCATGTTTCATTTAAAATAAAATGTTTCTC   1022
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AATTGAAACCTTAAGGCTGTACTTTAATTCTTCATGTTTCATTTAAAATAAAATGTTTCTC   1020

Qy   1023 ATTAACTCTGATGGAAAA   1040
          ||||||||||||||||||
Db   1021 ATTAACTCTGATGGAAAA   1038
```

RESULT 2

ABL61981

```
ID   ABL61981 standard; DNA; 1172 BP.
XX
AC   ABL61981;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Colon adenocarcinoma related gene sequence SEQ ID NO:318.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   22-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235377P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   28-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0261111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0268891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237178P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
```

PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PS Claim 1; SEQ ID NO 318; 44pp; English.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABI61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (II) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;
XX

Query Match   41.4%; Score 431; DB 6; Length 1172;
Best Local Similarity   70.9%;
Matches 660; Conservative   0; Mismatches 255; Indels   16; Gaps   6;

```
Qy    66 GAGTTGGTCACAATGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAGAAATTCACGAG 125
         ||||  ||| ||||||||||||||||||||| ||||||| |||| || ||||||||||||||
Db   106 GAGCTGGCCCATAATGGCAGGTGAAGAAATTAATGAAGACTATATCCAGTAGAGAAATTCACGAG 165

Qy   126 TCTTTAACAGCCCTGGAGAGCTCCTGGGGTGCTGGTGGACCACATGCTGAGAGACCATGATG 185
         | |||  |||| | |||||| ||   ||| | |||||||   |||||||||||||||||||||
Db   166 TATTTGTCAGCGTTTGAGAATTCCATTGGTGTCTGTGGATGAGATGCTGAAGACCATGATG 225

Qy   186 GCTGCTTTCTTAGAAACGAGAGTTGTTGCGACCCAATTGGAACACGAACAAAGGTGGGAT 245
         |||| ||| || |||| ||||||||||||||   ||  || ||||| ||||| |||||||
Db   226 TCTGTTCTAGAAAATGAGAGTTGTTGCGAGACCCACTGTGAACAAGCAAAAACTGGGAT 285

Qy   246 TTAGTTTCTGCATACACCTTAAATTCAAATGTTTTATTTGGCAACTCAAGGAGTT 305
         || |||||||||||||| ||||||||||||||||| |||||||||| |||||||
Db   286 TTGGTTTCTGCATACACATTAAATTCAATGTTTTTGGTTTTATTTGGCAACCCAAGGAGTT 345

Qy   306 AATCCCAAAGAGCATCCAGTGAAGCAGGACTGGAAAGAATCAGAGTTCTTACATGAACAGA 365
         ||||| || || ||||||||||  |  || ||||  |||||||||| | | |||||||| |
Db   346 AATCCTAAGGAACATCCAGTGAAATACAGGAATTCAGAGTATATATGAACACGA 405

Qy   366 GTTAAAGAAAATAACAGACAGAAGAGCTGCCAAGCACAGAGGTTGCTGCTTCGAGA 425
         || || ||||||||||||||| ||| |||  ||||||||||||||  |||||| ||
Db   406 GTCAAGGAAATAACAGACAGAGAGGTGCAGCTTCAAGA 465

Qy   426 TTTGTCAAGAGAGGCACTCTGGGAACCCAAACGAAAAAGCACCAAAAGTGGGCTAATAAA 485
         |||||  || |  ||  ||||||||| ||||  |||||| | || ||||||| ||||||||
Db   466 TTTGTAAAAATGCCCTCTGGGAACCAAAAATCGAAAAAAATGCATACAAAAAAGTTGCCAATAAA 525

Qy   486 GGGAAAAAGCAAAACACTAATCTTTTTGGTTTGATGTACATGTTTTCAAAAAAGTACA-TCCT 544
         ||||||||| ||||| ||||| ||||||||| ||||| |||| ||||||||| ||||| ||||
Db   526 GGAAAAAGTAAAAGTAAACTTTTTGGTTTTGATGTGTACACATATTCAAAAAGTACATTAAT 585
```

```
Qy   545 TTTTAATCAGTTTACAATGTAGTTATGTGACCATGTGGTGGTTTAAAATGGATTCCTTTTGG 604
         ||||||||||||||||||||| |||||||| | |||| || ||||||||||||  ||||||
Db   586 ATGTAATCACAGTAAATATGTAAAGCTAAATACTTCCTCTCCAAAGATCATTATCTTTATT 645

Qy   605 AATTCAATGTATAAAATTTACACAATTACACATTCATTTGTGATACTGGAATCTTTTTTTCTGAGAAA 664
         ||||||||| |||||||||||||||||||||||||||| |||||||||  ||||||  | |||||||
Db   646 GAT-----TAGCACTGAGGGATTTTAACATTGTGGATATATATTATATAATTATAATTACCAT 700

Qy   665 GATTAAGTTGTCTTGTGTTGATTTTCATAAAAGCATCATGATGTGTTAATATTGTAAGA 724
         |||||||| ||  ||||| |||||||||||||| || ||||| |||||||||||| |||
Db   701 CTCTTGATGAGACTGTTAATTCTTTATATAGGTCAGTCTGTAAACATTTGTAGTGTTTGAAAT 760

Qy   725 TATTCTTATAAGCAGTTGTGAAATCCAAAATGTTTCTCGTAAACATTTGTAGTGTTTGAAAT 784
         |||| |||||||  ||||||||||||||||| ||||||| ||||||||| ||||||||||||
Db   761 AGCTGTGAAA------TTTAAGTGAAAATGTTCTTTGTAAACATTTGCTACTATTTTAAAT 813

Qy   785 GAACAATGATATTATTGAAAGTGTGCTATTCGTTGAGACCTCGGAGGTGTGAAGGACATTTGTTTT 844
         ||| ||||| ||||| ||||  ||| ||| ||  ||  | | | |  | |||| |||||||||||
Db   814 GAATAATGACCTTATGAAGTATGCTATCTGTAG-GCTGAAATTATAGGTACAATCTGTTTT 872

Qy   845 CAGTAATGCATGAGAAATACAGTGACTTAAATACCCACTCTGTTTCTCGTTTCAGTTAGTTCA 904
         ||  |||| | || || |||| ||| |||||| |  | |||||| ||  ||||| |||| |||
Db   873 CACTATATGATTATTAAGAAAAGCG-TGAAATGACTTAAAATGTTCATTTTCTGTATTAGA 931

Qy   905 ACATGTTTCGTGCATTTTTTTTTTTTTTTGAGTTAATTCTGTCTGATTATTCAAAGT-CAAA 963
         ||||||| | |||| |||||||||||||| |||||| |||  ||||||||||||||| |||
Db   932 TACTTTATCATGTTTTCATGATTTTAGGAATTACTCGCTTTGTTGATATTCAAAGTGTGAA 991

Qy   964 ATTGAAACCTTAAGGCGTGTACTTTTAATTCTT 994
         | | ||| ||||| | ||||||||| |||||||
Db   992 ACTAAAAGTTTATNGGTTGTACTTTAATTCTT 1022
```

ID ABK84717
ID ABK84717 standard; cDNA; 1172 BP.
XX
AC ABK84717;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1288.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
OS Homo sapiens.
XX
PN WO20022899-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.

XX
PS
XX
CC   Claim 1; SEQ ID NO 1288; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) identified by
CC   DNA chip analysis as given in the specification, and comparing by
CC   differential expression of Gs is indicative of GCA, where
CC   expression level to an expression level in an unactivated GC, where
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GCA preferably in an
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;

Query Match   41.4%; Score 431; DB 6; Length 1172;
Best Local Similarity   70.9%;
Matches 660; Conservative   0; Mismatches 255; Indels   16; Gaps   6;

Qy   66 GAGTTGGTCACAATGGCAGGTGAAGAAGATTATNCCCGTAGAAAATTCACGGAG 125
             ||||| |||||||||||||||||||||||||||| ||||||||| |||||||||
Db   106 GAGCTGGCCATAATGGCAGGTGAAGAAATTAATGAAGACTATATCCAGTAGAAAATTCACGGAG 165

Qy   126 TCTTTAACAGCCCTGGAGGAGCTCCCTGGGGTGCTGTTGGACACATCGTCGAAGACCATGATG 185
             |||||||||| ||||||||||||||| |||||| ||||||| ||||||||||||||||
Db   166 TATTTGTGTCAGCGGTTTGAGAAATTCCATTGGTGTCGCTGTGGGATGAGATGCCTGAAGACCATGAG 225

Qy   186 GCTGTGTTTCTAGAAACCAGAGTTGCTTTGCGACCCAATTGGAGACAAGCAAAGGTGGAT 245
             |||||||||||||||||||||||||| |||||||||||||||||||
Db   226 TCTGTTTCTAGAAATGAGTTGTTGCGAGAGTTGCAATCACTTGAACAAGCAAAGTGGAT 285

Qy   246 TTAGTTTCTGCATACACCTTAAATTCAAGTGTTTTGGGTTTATTGGCAACTCAAGGAGTT 305
             ||||||| ||||||||||| ||||||||||||||||||||||||||||||||
Db   286 TTGGTTTCTGCATACACAATTAAATCAAGTGTTTTGGGTTTATTGGCAACCCAAGGAGTT 345

Qy   306 AATCCCAAAGAGACATCCAGTCAAGCTGGGAAAGAATCAGAGTCTACATGAACAGA 365
             |||||||| |||||||||||||||||||||||||||||||||||||
Db   346 AATCCTAAGGAACATCCAGTTAAAACAGAGAATTCAGAGTTATATGAACAGA 405

Qy   366 GTTAAAGAAATAACAGAAGGCTGCCAAGCTGGACAAGAGAAGGCTGCAGAGTGTGCTGAGA 425
             ||| |||||||||||||||||||||||||||||||||||||||||||||
Db   406 GTCAAGGAAATAACAGACTGGAAAGCACACCAAAAGAGTTGCCAATAA 425

Qy   426 TTTGTCAAGAGGCCACTCTGGGAACCCAACCTGCGCCTTGGGAACTGGCTTAATAAA 485
             ||| |||||||||||||||||||||||||||||||||||||||||| ||
Db   426 GTCAAGGAAATAACAGACTGGAAAGCACACCAAAAGAGTTGCCAATAA 485

Qy   466 TTTGTAAAAAATGCCCTCTGGGAACCAAAATCCAAAAAATGCATAAAAGTTGCCAATAA 525
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db   466 TTTGTAAAAAATGCCCTCTGGGAACCAAAATCCAAAAAATGCATAAAAGTTGCCAATAA 525

Qy   486 GGGAAAAGCAAAACACTAACTTTTGGTTTGGTTTTGATTGACATAGTTTTCAAAAAGTACA 544
             ||||| |||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   526 GGAAAAAGTAAAAAGTTAACTTTTGTTGGTTTTGATTGACATAGTTTTCAAAAAGTACATAAT 585

Qy   545 TTTTAATCAGTTTACAATGGTAGTTATGTGACCATGGTGTTTAAATGGATTCCTTTGG 604
             |||| |||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   586 ATGTAATCACAGTAATATGTAAAGCTAAATACTTCTCTCCAAAAGATTCATTATCTTTATT 645

Qy   605 AATTCATGTTATAAATTTACACATTACATTCATTGTGATACTGAATCTTTTTTTTGCTGAGAA 664
             ||||||||| |||| |||| |||| || ||| |||
Db   646 GAT-----TAGCACTGAGGAGTTTTAACCATTGTGATATTATATTTTTTTAAATTTATACCAT 700

Qy   665 GATTAAGGTTGTCTTTGTTGGATTTTCATATAAAGCCATCATATAGTTGTTTTAAATTGTAAGA 724
             |||||||||||||||||||| ||||| |||||||||||| ||||||| |||||||||| ||||
Db   701 CTCTTGATGAGACTCTTATTTCTTTATATAGGTCAGTCTTGCAAGTAGCATTTTTAAAGC 760

Qy   725 TATTCTATAAGCAGTTGTGAAATCCAAAATGTTCTCTGTAAACATTTGTTAGTGTTTGAAAT 784
             ||||||||||||||||||||| |||||||||||||||||||||||||| |||||||||||
Db   761 AGCTGTGAAA-------TTTAAGTGAAAGTTCTTTGTGAAACATTTGTACTATTTTAAAT 813

Qy   785 GAACAATGATATATGAAGGTGTGCATCGTAGACCTCGAGGGTGTAAGGACATTTGTTTT 844
             |||||||| ||||| |||||| |||||||||||||||| |||||||||||||||||||
Db   814 GAATAATGACCTTATGAAGTAGCTAATCTGTAG-GCTGAAAATTATAGGTACATCTGTTTT 872

Qy   845 CAGTATGATCAGAGAAATACAGTTGACTTAAATACCCACACTCGTTCTTCAGTTAGTTCA 904
             ||||||| ||||||||||||| |||||||||||||||| ||||||||||||||||
Db   873 CACTATATGATATTAAGAAAGCG-TGAAATGACTTAAAAGTTCATTTTTTTCTGTGTATAGA 931

Qy   905 ACATGTTTCGTGATTTTTTTTTTTTTTTTTTGAGTAAGTAATTCTGTCTTGATATTCAAAGT-CAAA 963
             |||||||| |||||||||||||||||||||||||||||||||||||||||||||| ||
Db   932 TACTTTATCATGTTTTCATGATTTTTAGGAATTACTGCTTTGTTGATAATTCAAAGTGTGAA 991

Qy   964 ATTGAAAACCTTAAGGCTGCTGTACTTTAATTCTT 994
             |||||||||||| |||||||||||||||||||||||
Db   992 ACTAAAAGTTTATCGTGTTGGTACTTTAATTCTT 1022

RESULT 4
ADF81400
ID   ADF81400 standard; DNA; 1172 BP.
XX
AC   ADF81400;
XX
DT   26-FEB-2004 (first entry)
XX
DE   Leukaemia-related DNA sequence #1956.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
XX
OS   Unidentified.
XX
PN   WO2003039443-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.
XX
PR   05-NOV-2001; 2001EP-00126244.
PR   30-APR-2002; 2002EP-0009758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGS2ENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
PA   (SCHO/) SCHOCH C.
PA   (KERN/) KERN W.
XX
PI   Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI   Eils R, Brors B, Mergenthaler S;
XX
DR   WPI; 2003-505037/47.
XX
PT   Determining the subtype of leukemia cells and whether a patient sample
PT   contains leukemia cells or other cells, useful for treating leukemia,

Left column

```
PT  comprises determining the expression profile of a group of markers in a
PT  patient sample.
XX
PS  Disclosure; SEQ ID NO 1956; 2938pp; English.
XX
CC  The present invention relates to a method (M1) for determining the
CC  subtype of leukaemia cells and whether a patient sample contains
CC  leukaemia cells. The method comprises determining the expression profile
CC  of a group of markers in a patient sample. The method is useful for
CC  determining the presence of leukaemia cells, its types or subtypes, and
CC  for the preparation of a medicament for treating leukaemia.
XX
SQ  Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;

Query Match          41.4%;  Score 431;  DB 10;  Length 1172;
Best Local Similarity 70.9%;  Pred. No. 1e-89;
Matches  660;  Conservative   0;  Mismatches 255;  Indels  16;  Gaps  6;

Qy   66 GAGGTTGGTCACCAATGGCAGGTGAAGAAATGAATGAAGAGAATTATCCCGTAGAAAATTCACGAG 125
        ||||| |||  || ||||||||||||||||| || ||||||  ||||| |||||||||||||||||
Db  106 GAGCTGGGCCATAATGGCAGGTGAAGAAATTAATGAAGAACTATCCAGTAGAAAATTCACGAG 165

Qy  126 TCTTTAAACAGCCCCTGGAGAGCTCCCTGGGGTTGCTGTGGACGACATGCTGAAGACCATGATG 185
        | ||| || |  |   || ||| ||| ||| |||||||||| ||| ||| ||| |||||||||||
Db  166 TATTTGTCACGGTTTGAGAATTCCATTGGTGCTGTGGATGAGAGTTGGACCAAGGAGTT 225

Qy  186 GCTGTTTCTAGAAACGAGGTTGTTGTTGCAGAAGTTGGACCCATTGGAACGAAGCAAAGTTGGAT 245
        |||||||||||||| |||||||||||| ||| ||| |||| || | ||||| || ||||||||||
Db  226 TCTGTTTCTAGAAAATGAGTTGTTGGATCCACTGAAGAGATGCTGAAGACCTGAAGAAAGTGGAT 285

Qy  246 TTAGTTTCTGCATACACCTTTAAAATTCAATGTTTTGGGTTTTATTTGGCACTCAAGGAGTT 305
        || |||||||||||||| ||||||||||||||||||||||||||||||||||| |||||||||
Db  286 TTGGTTTCTGCATACACATTAAAATTCAATGTTTTGGGTTTTATTTGGCACCAAGGAGTT 345

Qy  306 AATCCCAAAGAGCATCCAGTGAAAGGAACCTGGAAAGAAATCAGAGTTCTACATGAACAGA 365
        ||||| |||||| ||||||| |||| |||| || |||||||||||||| | ||||||||||
Db  346 AATCCTAAGGAACATCCAGTAAAAACAGGAATTGGAAAGAAATCAGAGTATATATGAACAGA 405

Qy  366 GTTAAAGAAATAACAGACAAGAAGAAGAGGCTGCCAAGCTGACACGAGGTGCTGCTTCGAGA 425
        || || |||| ||||||||||||||| |||||| |||||||||| | ||||| ||||||| |||
Db  406 GTCAAGGAAAATAACAGACAAGAAGAAAGGCTGGCAAGCTGACAGGGTGCAGCTTCAAGA 465

Qy  426 TTTGTCAAGAAGAGGCACTCTGGGGAACCCAAAGCACACCAAAAGTGGCTAAATAAA 485
        |||||  ||  | | | ||||||| |||| |||| | |||||||||| || |||||||
Db  466 TTTGTAAAAATGCCCTCTGGGAACCAAAATCGAAAAATCCATCAAAAGTTGCCCAATAAA 525

Qy  486 GGGAAAAGCAAAACAACTAATCTTTGTGATGTACATGTTTTCAAAAAGTACA-TCCT 544
        || |||||  ||| | ||| |||||| || ||||| ||| ||||||||||| | |||
Db  526 GGAAAAAGTAAAAGTAACTTTTTTGGTTTGATGGTACACATATTCAAAAAGTACAATTAAT 585

Qy  545 TTTTAAATCAGTTTACAAATGTAGTTATGTGACCATGTGTGTTTAAAATGGATTCCTTTTCG 604
        ||||||||||||| |  | || |||||||| |  |||||  || ||||||||| |  |||||
Db  586 ATGTAATCACAGTAATATATGTTAAAGCTAAATACTTCCTCTCCAAAGATCATTATCTTTATT 645

Qy  605 AATTCAIGTATAAATTACACATTACACATTTGTAGAATCTTTTTTTTGTGCTGAGAAA 664
        ||  | | |||||||||||| |||||||| ||||| || |||||| || || ||||||
Db  646 GAT-----TAGCACTGAGGATTTTAACATTGTGATAATATATATTTAATTAATTTACCAT 700

Qy  665 GATTAAGTTGTCTTTGTTGGATTTTCATATAAAGCATCATGATGTGTTTAATATTGTAAGA 724
        || |||| |  |||| |||| | |||| ||||||||||| || || |||||||||| ||||
Db  701 CTCTTGATGAGACTCTTATTCTTTATTATAGGTCAGTCTTGCAAGTACCATTTTATAAGC 760

Qy  725 TATTCTATAAGCAGTTGTGTAAAATCCAAAGTTCTCTGTAAACAATTGTAGTGTTTTGAAAT 784
        ||||||||||||||| || ||||| ||||||| ||||||||||| |||||| ||| ||||||
Db  761 AGCTGTGAAA------TTTAAGTGAAAAGTTCTTTGTAAACATTTGTACTATTTTAAAT 813

Qy  785 GAACAATGATATATTGAAGTGTGTCTATCGTGTAGACCTCGAGGGTGTAAGGACACATTTGTTTT 844
        |||||||||| |||||||||| | ||||||||||| |  |||| || ||  ||||||| ||||||
Db  814 GAATAAATGACCTTATGAAGTATGCTATCGTGTAG-GCTGAAATTAATAGGTACATCTGTTTT 872

Qy  845 CAGTAATGATGAGAAATACAGTGACTAAAATACCCACTCTGTTTCTGTTCAGTTCA 904
        || |||||| || ||| |||||||| |||| | ||| ||||| | || || |||||
Db  873 CACTATATGATATTAAGAAAGCG-TGAAATGACTAAAAGTTCAITTTTTCTGTATAGA 931
```

Right column

```
Qy  905 ACATGTTTCGTTGATTTTTTTTTTTTTTGAGTAATTCGTGTCTTGATATTCAAAGT-CAAA 963
        ||||| ||  | |||||||||||||||||||||| || |||  |||||| |||||| ||||
Db  932 TACTTTATCAITGTTTTCATGAITTTAAGGAATTACTCGCTTTTGTCTTGGATAITTCAAAAGTGTGAA 991

Qy  964 ATTGAAACCTTAAGCTGTACTTTTATTCTT 994
        |||||| |||||| |||||||||| |||||
Db  992 ACTAAAAGTTTATGGTTGTACTTTAATTCTT 1022

RESULT 5
ID  ADF81399 standard; DNA; 1172 BP.
XX
AC  ADF81399;
XX
DT  26-FEB-2004  (first entry)
XX
DE  Leukaemia-related DNA sequence #1955.
XX
KW  Cytostatic; Gene therapy; leukaemia; ss.
XX
OS  Unidentified.
XX
PN  WO200303944 3-A2.
XX
PD  15-MAY-2003.
XX
PF  04-NOV-2002; 2002WO-EP012303.
XX
PR  30-APR-2002; 2002EP-00009758.
PR  05-NOV-2001; 2001EP-00126244.
XX
PA  (DEKR-)  DEUT KREBSFORSCHUNGSZENTRUM.
PA  (UYLU-)  UNIV LUDWIG MAXIMILIANS.
PA  (HAFE/)  HAFERLACH T.
PA  (SCHO/)  SCHOCH C.
PA  (KERN/)  KERN W.
XX
PI  Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI  Eils R, Brors B, Mergenthaler S;
XX
DR  WPI; 2003-505037/47.
XX
PT  The present invention relates to a method (M1) for determining the
PT  patient sample.
XX
CC  Determining the subtype of leukaemia cells and whether a patient sample
CC  contains leukaemia cells or other cells, useful for treating leukemia,
CC  comprises determining the expression profile of a group of markers in a
CC  of a group of markers in a patient sample. The method is useful for
CC  determining the presence of leukaemia cells, its types or subtypes, and
CC  for the preparation of a medicament for treating leukaemia.
XX
PS  Disclosure; SEQ ID NO 1955; 2938pp; English.
XX
SQ  Sequence 1172 BP; 403 A; 155 C; 237 G; 377 T; 0 U; 0 Other;

Query Match          41.4%;  Score 431;  DB 10;  Length 1172;
Best Local Similarity 70.9%;  Pred. No. 1e-89;
Matches  660;  Conservative   0;  Mismatches 255;  Indels  16;  Gaps  6;

Qy   66 GAGGTTGGTCACCAATGGCAGGTGAAGAAATGAATGAAGAGAATTATCCCGTAGAAAATTCACGAG 125
        ||||| |||  || ||||||||||||||||| || ||||||  ||||| |||||||||||||||||
Db  106 GAGCTGGGCCATAATGGCAGGTGAAGAAATTAATGAAGAACTATCCAGTAGAAAATTCACGAG 165

Qy  126 TCTTTAAACAGCCCCTGGAGAGCTCCCTGGGGTTGCTGTGGACGACATGCTGAAGACCATGATG 185
        | ||| || |  |   || ||| ||| ||| |||||||||| ||| ||| ||| |||||||||||
Db  166 TATTTGTCACGGTTTGAGAATTCCATTGGTGCTGTGGATGAGAGTTGGACCAAGGAGTT 185

Qy  186 GCTGTTTCTAGAAACGAGGTTGTTGTTGCAGAAGTTGGACCCATTGGAACGAAGCAAAGTTGGAT 245
        |||||||||||||| |||||||||||| ||| ||| |||| || | ||||| || ||||||||||
Db  226 CAGTTGGACCCATTGGAACGAAGCAAAGTGGAT 245
```

Key
FH   CDS
FT   Location/Qualifiers
FT   118..543
FT   /*tag= a

OS   Homo sapiens.

KW   C1D; human; apoptosis; tumour; gene therapy; treatment; ss.

DE   Human C1D cDNA.

DT   17-MAR-2000 (first entry)

AC   AAZ43927;

ID   AAZ43927 standard; cDNA; 1158 BP.

RESULT 6
AA243927

FT   /product= "C1D"

PN   DE19824811-A1.

PD   09-DEC-1999.

PF   03-JUN-1998; 98DE-01024811.

PR   03-JUN-1998; 98DE-01024811.

PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

PI   Rochbarth K, Stammer H, Werner D;

DR   WPI; 2000-063506/06.
DR   P-PSDB; AAY51024.

PT   Inducing apoptosis by overexpressing the C1D gene, particularly for
PT   treating tumors.

PS   Claim 3; Fig 1; 10pp; German.

CC   This invention describes a novel method for inducing apoptosis which
CC   comprises overexpressing the C1D gene (I). The method is particularly
CC   used to treat tumors and can also be used in gene therapy. The method has
CC   no side effects on normal cells (contrast known methods of inducing
CC   apoptosis such as cytotoxins and radiation), and may be effective on
CC   cells resistant to conventional treatments. Overexpression of (I) is
CC   sufficient itself to induce apoptosis but the effect may be increased
CC   when used in combination with other anti-tumour methods. When cells
CC   transfected with (I) undergo apoptosis, they release factors that kill
CC   neighboring, non-transfected cells (bystander effect). This sequence
CC   encodes the human C1D protein described in the method of the invention

SQ   Sequence 1158 BP; 388 A; 155 C; 237 G; 378 T; 0 U; 0 Other;

Query Match   41.3%; Score 430; DB 3; Length 1158;
Best Local Similarity 70.1%; Pred. No. 1.7e-89;
Matches 652; Conservative 0; Mismatches 265; Indels 13; Gaps 5;

Qy   66 GAGGTTGGTCACAATGGCAGGTGGAGAAATGAATGAAGATATATCCGTAGAAATTCACGAG 125
              ||||||| ||||||| ||||||||| ||| ||||| |||| |||||||||||||||||
Db   992 ACTAAAAGTTTATGGTTGTTGTACTTTAATTCTT 1022

Qy   106 GAGCTGGCCATAATGGCCAGTAGAGCTGGAAGAAAATTAATGAAGACTATCCAGTAGAAATTCACGAG 165
Db   964 ATTGAAACCTTAAGGCGTGCTACTTTAATTCTT 994

Qy   126 TCTTTAACAGACCCCTGGAGAGCTCCCTGGGGTGCTGTGGACCGACATGCTGAAGACCATGATG 185
Db   932 TACTTTAATCAGTGTTTCATGATTTTTGAAGAA 963

Qy   166 TATTTGTCAGCCTTTGAGAAATTCCATTGGTGCTGTGGATGAGATGCTGAAGACCATGATG 225
Db   905 ACATGTTTCGTGATTTTTTTTTTTTGAGTAAT 931

Qy   186 GCTGTTTCTTAGAAACCAGTGTTGTCAGAAGTTGGACCCATTGGAACAAGCAAAGGTGGAT 245
Db   873 CACTATAATGATTAATTAAGAAGTGATTTAAATCCCACTTCTGTTTCTGTATAGA 904

Qy   226 TCTGTTTTTCTTAGAAAACCAGTTGTGTTCAGAGTTGGATCCAACTTGAACAAGCAAAGAGAGTGGAT 285
Db   845 CAGTAAATGAGAGAAAATACAGTGACTTAAATACCCACTCTGTTCTCTGTTCTCAGTTAGTTCA 872

Qy   246 TTAGTTTCTGCATACACCTTAAATTCAATGTTTGGGTTATTGGCAACTCAAGGAGTT 305
Db   814 GAATAATGACCTTATGAAGTATGCTATCTGTAG-GCTGAAATTATAGGTACACTCGTTTT 844

Qy   286 TTGGTTTTCTGCATACACAATTAAATTCAATGTTTGGGTTATTGGCAACCCAAGGAGTT 345
Db   785 GAACAAATGATTATTTATGAAGTGTGCTACTTCGAGGTGTAAGGACAACATTGTTGTTTTGTTAAAT 813

Qy   306 AATCCCAAAGAGCATCCAGTGAAGCATCAGAGTTCAGAGTCTTACACGGACAGA 365
Db   761 AGCTGTGAA-------TTTAAGTGAAATGTTCTTTGTGAAGTAGAACATTTGTACTGATTTGTTTAAAT 784

Qy   346 AATCCTAAGGAACATCCACTAAAACAGAGAATTGGAAAGAAAATCAGAGTTAAGAACAGA 405
Db   725 TATTCTATATAAAGCAGTTGTGAAATCCAAAAGTTCTCTGTAAAACAAGTTGTAGTGTTGTTGAAAT 760

Qy   366 GTTAAAAGAAAATAACAGACAAGAAGAAGGCTGCCAAGCTGGACAGAGGGTGCTGCTTCGGAGA 425
Db   701 CTCTTGAATGAGACACTCTTATTTCTTTATATAGGTCAGTCTTGCAAGTACCATTTTATAGC 724

Qy   406 GTCAAGGAAAATAACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGGTGCAGCTTCAAGA 465
Db   665 GATTAAGTTCGTCTTTTGTTTTGTGAGTTTTCATATAAAGCATCATGGTGTTTTAATTATGTTAAGA 700

Qy   426 TTTGTCAAGGAAGGCCACTCTGGGACCAAAAGCAACACCAAAAAAGTGGCTAATAAA 485
Db   646 GAT-----TACCACTGAAGGATTTAACATTTGTGATAATCTGAAATCTTTTTTTGTGAGAAA 664

Qy   466 TTTGTAAAAAAATGCCCTCTGGGAACCAAAATCGAAAAATTGAAAAAAAGTTGCCAATAAA 525
Db   605 AATTCATGTATAAAATTTACACATTTGTGTATACATGAAATCTTTTTTTGCTGAGAAA 645

Qy   486 GGGAAAAGCAAACACTAATCTTTTGGTTTTTGATGTAACATAATTCCTTTTCG 544
Db   586 ATGTAATCACAGTAATATAATGTAAAGCTAAAATACTTCCTCCAAAGATCATTATCTTTATT 604

Qy   526 GGAAAAAGTAAAAAGTTAACTTTTTTGATGTACACATATTCAAAAAAGTACACATTAAT 585
Db   545 TTTTAATCAGTTTACAATGTAGTTAGTTGATGACCATGTGGTGTTAAATGGATTCCTTTTCG 585

```
Db   583 AATATGTAATCACAGTAATATGTAAAGCTAAATACTTCCTCTCCAAAGATCATTTATCTT  642
             |||                  |||                          |||
Qy   606 ATTCATGTATAAAATTTACACATTACATTGTGATCTGAATCTCTTTTTTGCTAGGAAAG   665
Db   643 TATTGATTACGACTGAGGATTTTAACATTGTGATTATATATATATTATATTACCATC    702
Qy   666 ATTAAGTTGTCTTTGTTTGATTTTCATATAAAAGCATCATGATGTGTTAATATTGTAAGAT 725
Db   703 TCTTGATGAGACTCTTATTTCTTTATATAGGTCAGTCGTTCGAAGTACCATTTATAAGCA   762
Qy   726 ATTTCTATAAGCAGTTGTGAAAATCCAAAATGTTCTCGTTAACATTTGTAGTGTTTGAAATG 785
Db   763 GCTGTGAAA-------TTTAAGTGAAAAGTTCTTTGTTAAACATTTGTACTATTTTAAATG   815
Qy   786 AACAATGATATTATTATGAAAGTGTGCTATTCGTAGACCTTCGAGGTGTGAAGGACATTTGTTTTC 845
Db   816 AATAATGACCTTATGAAGTATGCTTAACTCGTGAG-GCTGAAAATATATAGGTACATCTGTTTTC  874
Qy   846 AGTAATGATGAGAAATACAGTGACTTAAATACCCACTCTGTTTCGTTCAGTAGTTCAA    905
Db   875 ACTATATGATATTAAGAAAAGCG-TGAAAATGACTTAAATGTTCATTTTTCTGTATAGAT   933
Qy   906 CATGTTTCGTGATTTTTTTTTTGAGTAATCTCTGTCTTTGATATTCAAAGT-CAAAA     964
Db   934 ACTTTATCATGTTTTTCATGATTTTTAGGAATTACTCGCTTTGTTGATATTCAAAGTGTGAAA  993
Qy   965 TTGAAACCTTAAGGCTGTTACTTTAATTCTT                              994
Db   994 CTAAAAGTTTATGGTTGTACTTTAATTCTT                              1023
```

RESULT 7
ID   ADJ92841 standard; DNA; 426 BP.
XX
AC   ADJ92841;
XX
DT   06-MAY-2004 (first entry)
XX
DE   Human co-repressor SUN-CoR DNA.
XX
KW   Nuclear receptor; therapy; diabetes; hyperlipidaemia; obesity;
KW   atherosclerosis; human; co-repressor; gene; ds.
XX
OS   Homo sapiens.
XX
XX
PN   US2003228607-A1.
XX
PD   11-DEC-2003.
XX
PF   14-APR-2003; 2003US-00414692.
XX
PR   15-APR-2002; 2002US-0372650P.
XX
PA   (WAGN/) WAGNER B L.
PA   (SCHU/) SCHULMAN I G.
XX
PI   Wagner BL, Schulman IG;
XX
DR   WPI; 2004-167207/16.
XX
PS   Disclosure; SEQ ID NO 42; 99pp; English.
XX
CC   The invention relates to screening methods for identifying compounds that
CC   bind to nuclear receptor and exhibit cell type specific actions. The
CC   invention relates to modulators having an improved therapeutic profile.
CC   The method is useful for identifying compounds that bind to a nuclear
CC   receptor and exhibit cell type specific actions. It is also useful for
CC   identifying modulators of nuclear receptors that are useful in treating
CC   diseases e.g., diabetes, hyperlipidaemia, obesity, atherosclerosis, etc.
CC   The present sequence is human co-repressor DNA used to illustrate the
CC   method of the invention.
XX
SQ   Sequence 426 BP; 150 A; 76 C; 93 T; 107 G; 0 U; 0 Other;

Query Match   40.2%; Score 418; DB 12; Length 426;
Best Local Similarity 98.8%; Pred. No. 7, 5e-87;
Matches   421; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

```
Qy   78 ATGGCAGGTGAAGAAATGAAATGAATGAATAATCCCGTAGAAATTCACGAGTCTTTTCTAGA 137
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGGCAGGTGAAGAAAATGAATGAATATTCCCGTAGAAATTCACGAGTCTTTTCTAGA    60
Qy  138 CTGGAGAGCTCCCTGCGTGCTGTGCTGGACGACGAACATGCTCAAAGACCATGATTGGCTGTTCTTTCTAGA 197
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTGGAGAGCTCCCTGGGTGCCGTGGATGACAATGACGAAGGCTGTTTTCTAGA      120
Qy  198 AACGAGTTGTTGCGAGAGTTGGACCCATTGGAACACAAGCAAAGGTGGATTAGTTTTCTGCA 257
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AATGAGTTGTTGCAGAGTTCGGACCCATTGAACACAAGCAAAGGTGGAATTGAATTAGTTTTCTGCA 180
Qy  258 TACACCTTAAATTCAATGTTTGTGGGTTTTATTTGGGCAACTCAAGGAGTTAATCCCAAAGAG 317
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TACACCTTAAATTCAATGTTTGTGGGTTTTATTTGGGCAACTCAAGGAGTTAATCCCAAAGAG 240
Qy  318 CATCCAGTGAAGCAGCAGAACTCGGAAAAGTCAGAGTCTTACAGTGAACAGAGTTAAAGAAATA 377
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CATCCAGTGAAGCAGCAGAACTCGGAAAAGTCAGAGTCTTACAGTGAACAGAGTTAAAGAAATA 300
Qy  378 ACAGACAAGAAGAGAAGGCTGCCAAGCTGGACAGAGGTGCGCTTCGAGAGATTTGTCAAGAAG 437
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ACAGACAAGAAGAGAAGGCTGCCAAGCTGGACAGAGGTGCGCTTCGAGAGATTTGTCAAGAAC 360
Qy  438 GCACTCTGGGACCCAAAACGCACACCACCAGAGGGTGCTAATAAAGGGAAAAAGCAAA      497
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GCACTCTGGGACCCAAAACGCACACCACCAGAGGGTGCTAATAAAGGGAAAAAGCAAA      420
Qy  498 CACTAA 503
          ||||||
Db  421 CACTAA 426
```

RESULT 8
ID   ACH68585/c standard; DNA; 586 BP.
XX
AC   ACH68585;
XX
DT   29-JUL-2004 (first entry)
XX
DE   Human genome derived single exon probe #1780.
XX
KW   Human; probe; ss; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG, Rank DR, Hanzel DK;
XX

DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing events, for assessing genomic alterations or as tools for
PT   surveying tissues.
XX
PS   Claim 15; SEQ ID NO 1780; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately
CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   tissues to detect the presence of expressed messages that contain their
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in identifying and characterising
CC   alternative splicing events, in detecting and characterising gross
CC   smaller genomic alterations in the genomic locus that includes their exon,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030194704
XX
SQ   Sequence 586 BP; 145 A; 138 C; 109 G; 194 T; 0 U; 0 Other;

Query Match   37.8%; Score 392.6; DB 12; Length 586;
Best Local Similarity   84.1%; Pred. No. 6.3e-81;
Matches 443; Conservative   0; Mismatches   84; Indels   0; Gaps   0;

Qy   15 TGGCGTCATCATCGTGCGACCTATTTCCCGAGACAGGCCGTCCACGGTATTGAGTTGGTTC 74
Db   581 TGTCCGTTCATCCTTGCCCGACCCGCTTTCCGGAGGGCTGGACGTCGAAGCGCCGTG 522

Qy   75 ACAATGGCAGGTGAAGAAATGAATGAAGAATTATCCGTAGAAAATTCACGAGTCTTTAACA 134
Db   521 ATAACGGCAGGTGAAGAAATTAATGAAGACTATCCAGTAGAAAATTCACCAGTATTTGTCA 462

Qy   135 GCCCCTGGAGGAGCTCCCTGGGTGTCGTGGACGACGCATGTCCAGTAGGCTTTTTTTTTCT 194
Db   461 GCCGTTGAGAATTCCATTGGTGCTGTGGAATGACCACGAGTCTGTTTTCT 402

Qy   195 AGAAACGACGTTGTTGTTGCAGAGTTGGAACAAGCAAAGGTGGATTTAGTTTCT 254
Db   401 AGAAATGAGTTGGTTCGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTTGGGTTTCT 342

Qy   255 GCATACACCTTAAATTCAATGTTTTTGGGTTTTATTTGGGTTAATCCCAAA 314
Db   341 GCATACACATTAAATTCAATGTTTTGGGTTTATTTGGGTTAATCCTAAG 282

Qy   315 GAGCATCCAGTGAAGCAGGAACTGGAAAGAATCTACATGACGAGAGTTAAAGA 374
Db   281 GAACAATCCAGTAAAACAGGAATTGGAAAGATCAGAGTGTATATGAACAGAGAGTCAAGGAA 222

Qy   375 ATAACAGACAAGAGAGAAGACAGGCTCGGCCAAGCTGGACGAGAGAGGTGTGCTTCGGAGATTTGTCAAG 434
Db   221 ATAACAGACAAGAGAAGACTGGCACAGGAGGTGCAGCTTCAAGATTTGTAAAA 162

Qy   435 AAGGCACTCTGGGGAACCCAAACGAAAAAAGCACACCAAAAAGTGGCTAATAAAAGGGGAAAGC 494
Db   161 AATGCCCTTCTGGGACCAAAAACCGAAAAAAGTGCAGAAAAACGGGAAAAGT 102

Qy   495 AAACACTAAATCTTTTGTTTTTCATGTACATGTTTTCAAAAAGTACAT 541
Db   101 AAAAGTTAACTTTTTTGGTTGTGTACACATATTCAAAAAAGTACAT 55

Db   281 GAACAATCCAGTAAAACAGGAATTGGAAAGATCAGAGTGTATATGAACAGAGAGTCAAGGAA 222
XX
PS   Claim 1; Fig 2; 68pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of two
CC   mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC   CF8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CF7
CC   and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC   involved in cellular functions, such as: regulation of metabolism and
CC   cell homeostasis, cell proliferation and differentiation, pathological
CC   cellular aberrations, or cellular defence mechanisms. The present cDNA
CC   sequence represents the reverse complement of the mammalian nuclear
CC   receptor cofactor CF7 gene sequence
XX
SQ   Sequence 426 BP; 105 A; 99 C; 64 G; 158 T; 0 U; 0 Other;

Query Match   31.4%; Score 326.8; DB 6; Length 426;
Best Local Similarity   85.4%; Pred. No. 9.5e-66;
Matches 364; Conservative   0; Mismatches   62; Indels   0; Gaps   0;

Qy   78 ATGGCAGGTGAAGAAATGAATGAAGAATTATCCGTAGAAAATTCACGAGTCTTTAACAGCC 137
Db   426 ATGGCAGCTGAAGAAATTAATGAAGGACTATCCAGTAGAAAATTCACCGATTATTGTCAGCA 367

RESULT 9
AAL43974/c
ID   AAL43974 standard; cDNA; 426 BP.
XX
AC   AAL43974;
XX
DT   27-SEP-2002 (first entry)
XX
DE   Mammalian nuclear receptor cofactor CF7 reverse complement sequence.
XX
KW   Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;
KW   metabolism regulation; cell homeostasis; cell proliferation;
KW   differentiation; pathological cellular aberration;
KW   cellular defence mechanism.
XX
OS   Mammalia.
XX
PN   WO20024233222-A2.
XX
PD   30-MAY-2002.
XX
PF   21-NOV-2001; 2001WO-EP013548.
XX
PR   21-NOV-2000; 2000EP-00125524.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Jackson D; Casari G, Suckow J;
XX
DR   WPI; 2002-566559/60.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.

Left column

```
Qy  138 CTGGAGAGCTCCCTGGGTGCTGTGTGTGACGACATGCTGAAGACCATGATGGCTGTTTCTAGA  197
Db  366 TTTGCGAATTCCATTGATGCTGTGTGGATGAGATGCTGAAGAACATGTCTGTTTTCTAGA     307
Qy  198 AACGAGTTGTTGCAGAAGTTGGACCCACTGGAATGGACTGAAGAACATGTCTGTTTTCTGCA   257
Db  306 AATGAGTTGTTGCAGAAGTTGGACCCACTGGAATGGACTGAAGAACATGTCTGTTTTCTGCA   247
Qy  258 TACACCTTAAATTCAAATGTTTTGGGTTTATTTGGGCAACTCAAGGAGTTAATCCCAAAGAG   317
Db  246 TACACCTTAAATTCAAATGTTTTGGGTTTATTTGGGCAACTCAAGGAGTTAATCCTAAGGAA   187
Qy  318 CATCCAGTAAAGCAGGAACTGGAAAGAATCAGAGTCTACATGACAGAGTTAAAGAAATA     377
Db  186 CATCCAGTAAAGCAGGAACTGGAAAGAATCAGAGTGTATATGACAGAGTCAAGGAAATA      127
Qy  378 ACAGACAAGAAGAAGGCTGCCAAGCTGGACAGAGGTGCGTCGTTCTTCGAGATTTGTCAAGAAG  437
Db  126 ACAGACAAGAAAAAGGCTGGACAGAGGTGCAGCTTCAAGATTTGTGTAAGAAAT            67
Qy  438 GCACTCTGGGAACCCAAACCAAAAAGTGACACACCAAAAGTGGGAAAAGCAAA            497
Db   66 GCCCCTCTGGGAACCAAAACCGAAAAAATACATCCAAAAGTTGCCCATAAAAGGAAAAGTAAA    7
Qy  498 CACTAA  503
Db    6 AGTTAA  1
```



```
RESULT 10
ID   AAL43973 standard; cDNA; 426 BP.
XX
AC   AAL43973;
XX
DT   27-SEP-2002 (first entry)
XX
DE   Mammalian nuclear receptor cofactor CF7 coding sequence.
XX
KW   Mammalian; gene; ss; nuclear receptor cofactor; CF7; CF8;
KW   metabolism regulation; cell homeostasis; cell proliferation;
KW   differentiation; pathological cellular aberration;
KW   cellular defence mechanism.
XX
OS   Mammalia.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..426
FT                  /*tag= a
FT                  /product= "Mammalian CF7 protein"
XX
PN   WO20024322-A2.
XX
PD   30-MAY-2002.
XX
PF   21-NOV-2001; 2001WO-EP013548.
PR   21-NOV-2000; 2000EP-00125524.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Jackson D, Casari G, Suckow J;
XX
DR   WPI; 2002-566559/60.
DR   P-PSDB; AA015405.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.
XX
PS   Claim 1; Fig 1; 68pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of two
```

Right column

```
CC   mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC   CF8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CF7
CC   and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC   involved in cellular function, such as: regulation of metabolism and
CC   cell homeostasis, cell proliferation and differentiation, pathological
CC   cellular aberrations, or cellular defence mechanisms. The present cDNA
CC   sequence encodes the mammalian nuclear receptor cofactor CF7 protein
XX
SQ   Sequence 426 BP; 158 A; 64 C; 99 G; 105 T; 0 U; 0 Other;

Query Match          31.4%;  Score 326.8; DB 6;  Length 426;
Best Local Similarity 85.4%;  Pred. No. 9.5e-66;
Matches  364;  Conservative   0;  Mismatches  62;  Indels   0;  Gaps   0;
```

```
Qy   78 ATGGCAGGTGAAGAAAATGAATGAAGATTATCCGTAGAAAATTCACGAGTCTTTACAGCC   137
Db    1 ATGGCAGCTGAAGAAAATTAATGAAGGACTATCCAGTAGAAATTCACGATTATTGTCAGCA   60
Qy  138 CTGGAGAGCTCCCTGGGTGCTGTGTGTGACGACATGCTGAAGACCATGATGGCTGTTTCTAGA  197
Db   61 TTTGCGAATTCCATTGATGCTGTGTGGATGAGATGCTGAAGAACATGTCTGTTTTCTAGA     120
Qy  198 AACGAGTTGTTGCAGAAGTTGGACCCACTGGAATGGACTGAAGAACATGTCTGTTTTCTGCA   257
Db  121 AATGAGTTGTTGCAGAAGTTGGACCCACTGGAACCAAGCAAGCAAAAGTGGAATTTGGGTTT   180
Qy  258 TACACCTTAAATTCAAATGTTTTGGGTTTATTTGGGCAACTCAAGGAGTTAATCCCAAAGAG   317
Db  181 TACACATTAAATTCAATGTTTTTGGGGTTTATTTGGGCAACTCAAGGAGTGAATCCTAAGGAA  240
Qy  318 CATCCAGTAAAGCAGGAACTGGAAAGAATCAGAGTCTACATGACAGAGTTAAAGAAATA     377
Db  241 CATCCAGTAAAGCAGGAATTGGAAAGAATCAGAGTGTATATGACAGAGTCAAGGAAATA      300
Qy  378 ACAGACAAGAAGAAGGCTGCCAAGCTGGACAGAGGTGCGTCGTTCTTCGAGATTTGTCAAGAAG  437
Db  301 ACAGACAAGAAAAAGGCTGGACAGAGGTGCAGCTTCAAGATTTGTGTAAGAAAT            360
Qy  438 GCACTCTGGGAACCCAAACCAAAAAGTGACACACCAAAAGTGGGAAAAGCAAA            497
Db  361 GCCCCTCTGGGAACCAAAACCGAAAAAATACATCCAAAAGTTGCCCATAAAAGGAAAAGTAAA   420
Qy  498 CACTAA  503
Db  421 AGTTAA  426
```



```
RESULT 11
ID   ADI28090
XX
AC   ADI28090;
XX
ID   ADI28090 standard; DNA; 420 BP.
XX
AC   ADI28090;
XX
DT   22-APR-2004 (first entry)
XX
DE   Human nuclear receptor cofactor CF6 sense coding strand.
XX
KW   gene; nuclear receptor cofactor; CF6; cellular function inhibition;
KW   metabolic disorder; immunological indication; hormonal dysfunction;
KW   neurosystemic disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..420
FT                  /*tag= a
XX
PN   WO200224728-A2.
XX
PD   28-MAR-2002.
XX
PF   17-SEP-2001; 2001WO-EP010744.
```

XX
PR 22-SEP-2000; 2000EP-00120722.
PA (LION-) LION BIOSCIENCE AG.
XX
PI Casari G, Jackson D;
XX
DR WPI; 2002-383179/41.
DR P-PSDB; ADI28083.
XX
PT Polynucleotide and polypeptide of novel nuclear receptor cofactor useful for screening drugs regulating cofactor-associated physiological responses e.g. hormonal dysfunctions.
XX
PS Claim 1; SEQ ID NO 1; 97pp; English.

CC The invention relates to an isolated polynucleotide encoding a nuclear receptor cofactor, also known as CF6. The polynucleotide or encoded protein is useful for construction of multiple nuclear receptor cofactor specific sequence alignments, preferably for the construction of protein sequence alignments. The protein is useful for screening agents capable of inhibiting the cellular function of the cofactor CF6. The polynucleotide is useful for making vectors and for transforming cells, both of which are ultimately useful for production of the CF6 protein. They are also useful as scientific research tools of the cofactor gene, e.g. to identify diseased or otherwise abnormal states. They are particularly useful for determining expression levels of the cofactor gene, e.g. to identify diseased or otherwise abnormal states. They are particularly useful for diagnostic purposes to e.g., identify deleted or mutant CF6 genes; or their measure expression. They are useful for developing inhibiting expression of the cofactor gene to determine physiological responses. The protein is useful for screening drugs for agonist and antagonist activity, and therefore, for screening for drugs useful in regulating physiological responses associated with the cofactors such as metabolic disorders, immunological indications, hormonal dysfunction, neurosystemic diseases. The proteins are also useful for developing antibodies for detection of the proteins. The polynucleotide can be used to design primers for a polymerase chain reaction and are also used to model the three-dimensional structure of the protein. This sequence represents the sense strand of the protein. (Note: this sequence appears to be a fragment of the CF6 gene as the sequence differs from the CF6 gene sequence given in Fig 1 (also designated SEQ ID NO: 1) of the specification.
XX
SQ Sequence 420 BP; 155 A; 64 C; 98 G; 103 T; 0 U; 0 Other;

Query Match   31.3%; Score 325.6; DB 6; Length 420;
Best Local Similarity 86.0%; Pred. No. 1.8e-65;
Matches 361; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy   78 ATGGCAGGCTGAGAAGAAATGAAGAGAGATTATCCGTAGAAGAGGTCTTTAAACAGCC 137
         ||||||| ||||||||||||  |||||||| ||||||||||||   | ||||||  ||||
Db    1 ATGGCAGCTGAGAAGAAATTAATGAGAGACTATCCGTAGAAATTCACGAGTCTTTCTAGA 60

Qy  138 CTGGAGGAGCTCCCTGGGTGCTGTGGACGACCATGGTAGATGGCTGTTTTCTTCAGCA 197
         ||||||||||  || ||| ||||||||| |||| ||| |||| |||||||||| || |
Db   61 TTTGGCGAAATTCCATTGGATGCTGTGGATGAGATGTCGTTTTCTTCTAGA 120

Qy  198 AACGAGTTGTTGCAGAAGTGGGACCCATTGGACAAAGCAAAGCTGGGATTTAGTTCTGCA 257
         ||||||||||||| ||| |||| ||||||||| |||||||| ||||||| ||||||||||
Db  121 AATGAGTTGTTGCAGAAGTTGGACCCACTTGAACAGCAAAAGTGGAATTTGGTTTCTCA 180

Qy  258 TACACCTTAAATTCAATGTTTTGGTTTTATTTGGCACTCAAGGAGTTAATCCAAGAG 317
         |||||||||||||||||   ||||||| |||||||| || |||| ||| || |||
Db  181 TACAATTAAATTCAAATGTTTTGGTTTTATTTGGCAACTCAAGGAGTGAATCCTAAGGAA 240

Qy  318 CATCCAGTTGAAGCAGAAGAACTGGAAGAATCAGAGTCTTACATGAACAGAGAGTTAAAGAAATA 377
         ||||||||| ||||||||| |||| |||||| |||||||| || ||||| ||||||  |||
Db  241 CATCCAGTTAAAGCAAGAAGAATTCAGGAATCAGAGTCACAGAGTGAAAGAAAT 300

Qy  378 ACAGACAAGAAGAAGGCTGCCAAGCTGGACACGAGAGGTGCTGCTTCGGAGATTTGTCAAGAAG 437
         ||||||||||||||||  | |||||||| |||| |||| |||| ||  ||||  ||||||||||
Db  301 ACAGACAAGAAAAAGGCTGGCAAGCTGGACAGAGGTGCGCAAGCTTCAAGAGATTTGTGTAAGAAAT 360
```

XX

```
Qy  438 GCACTCTGGGAACCCAAACGAAAAGCACACCAAAAAGTGGGCTAATTAAAGGAGAAAGCAAA 497
         || |  ||||||||||||  |||||  | || ||||| |||| | ||| |||||||||| |||
Db  361 GCCCTTCTGGGAACCCAAAACCGAAAAATACATCCAAAAGTTGCCCATAAAGGAAAAAGTAAA 420
```

RESULT 12
ID ADI28091/c
AC ADI28091;
XX
AC ADI28091 standard; DNA; 420 BP.
XX
DT 22-APR-2004 (first entry)
XX
DE Human nuclear receptor cofactor CF6 coding sequence complement strand.
XX
KW gene; nuclear receptor cofactor; CF6; cellular function inhibition;
KW metabolic disorder; immunological indication; hormonal dysfunction;
KW neurosystemic disease; ss.
XX
OS Homo sapiens.
XX
PN WO20224728-A2.
XX
PD 28-MAR-2002.
XX
PF 17-SEP-2001; 2001WO-EP010744.
XX
PR 22-SEP-2000; 2000EP-00120722.
XX
PA (LION-) LION BIOSCIENCE AG.
XX
PI Casari G, Jackson D;
XX
DR WPI; 2002-383179/41.
XX
PT Polynucleotide and polypeptide of novel nuclear receptor cofactor useful for screening drugs regulating cofactor-associated physiological responses e.g. hormonal dysfunctions.
XX
PS Claim 1; SEQ ID NO 2; 97pp; English.
XX
CC The invention relates to an isolated polynucleotide encoding a nuclear receptor cofactor, also known as CF6. The polynucleotide or encoded protein is useful for construction of multiple nuclear receptor cofactor specific sequence alignments, preferably for the construction of protein sequence alignments. The protein is useful for screening agents capable of inhibiting the cellular function of the cofactor CF6. The polynucleotide is useful for making vectors and for transforming cells, both of which are ultimately useful for production of the CF6 protein. They are also useful as scientific research tools for developing nucleic acid probes for determining expression levels of the cofactor gene, e.g. to identify diseased or otherwise abnormal states. They are particularly useful for diagnostic purposes to e.g., identify deleted or mutant CF6 genes; or their measure expression. They are useful for developing analytical tools such as antisense oligonucleotide for selectively inhibiting expression of the cofactor gene to determine physiological responses. The protein is useful for screening drugs for agonist and antagonist activity, and therefore, for screening for drugs useful in regulating physiological responses associated with the cofactors such as metabolic disorders, immunological indications, hormonal dysfunction, neurosystemic diseases. The proteins are also useful for developing antibodies for detection of the proteins. The polynucleotide can be used to design primers for a polymerase chain reaction and are also used to model the three-dimensional structure of the protein. This sequence represents the complement of the CF6 gene sense coding region.
XX
SQ Sequence 420 BP; 103 A; 98 C; 64 G; 155 T; 0 U; 0 Other;

Query Match   31.3%; Score 325.6; DB 6; Length 420;
Best Local Similarity 86.0%; Pred. No. 1.8e-65;
Matches 361; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy　78　ATGGGCGAGGTGAAGAAATGAAATGAAGAGATTATCCCGTAGAAATTCAGCGAGTCTTTAACAGCC　137
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　420　ATGGGCGAGGTGAAGAAATTAATGAAGAGCATATCCAGTAGAAATTCAGCGAGTTATTTGTCAGCA　361

Qy　138　CTGGAGAGCTCCCTGGGGTGCTGGTGACAGCACTATCCAGTAGAAGACCTATTCGGCTGGTTTTTTAGA　197
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　360　TTTCGGAATTCCATTGATGCTGTGTGATGAGATGCTGAAGAACACATGATGTCTGTTTCTAGA　301

Qy　198　AACGAGTTGTTGCAGAAGTTGGACAAGCAAAGGTGGGATTTAGTTTCTGCA　257
　　　||||||||||||||||||||||||||||||||||||||||||||||||||
Db　300　AATGAGTTGTTGCAGAAGTTGGGACCCACTTGAACACAAGCAAAAGTGGAATTGGGTTTCTGCA　241

Qy　258　TACACCTTAAATTCAATGTTTTTGGGTTTTATTTGGCAACTCAAGGAGTTAATCCCAAAGAG　317
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　240　TACACATTAAATTCAATGTTTTTGGGTTTATTTGGCAACTCAAGGAGTGAATCCTAAGGAA　181

Qy　318　CATCCAGTGAAGCAGGAACTGGAAAAGAATCTACATGAACAGAGTTAAAGAAATA　377
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　180　CATCCAGTAAAGCAGGAAATTGGAAGAAATCAGAGTATATATATGAACAGAGTCAAGGAAATA　121

Qy　378　ACAGACAAGAAGAAGAGGCTGCCAAGCTGGACAGAGGGTGCTGCTTCGAGATTTGTCAAGAAG　437
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　120　ACAGACAAGAAAAAAGGCTGGCAAGCTGGACAGAGGGTGCAGCTTCAAGAGATTTGTAAGAAAT　61

Qy　438　GCACTCTGGGAACCCAAACACAAAAGACACACCAAAAAGTGCTAATTAAAAAGGGAAAAAGCAAA　497
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　60　GCCCTCTGGGAACCAAAAACCGAAAAAATACATCCAAAGTTGCCCTATAAAGGAAAAAAGTAAA　1

RESULT 13
AAL43975
ID AAL43975 standard; cDNA; 420 BP.
XX
AC AAL43975;
XX
DT 27-SEP-2002 (first entry)
XX
DE Mammalian nuclear receptor cofactor CF8 coding sequence.
XX
KW Mammalian; gene; 88; nuclear receptor cofactor; CF7; CF8;
KW metabolism regulation; cell homeostasis; cell proliferation;
KW differentiation; pathological cellular aberration;
KW cellular defence mechanism.
XX
OS Mammalia.
XX
FH Key　　　　　Location/Qualifiers
FT CDS　　　　　1..420
FT　　　　　　　/tag= a
FT　　　　　　　/partial
FT　　　　　　　/product= "Mammalian CF8 protein"
FT　　　　　　　/note= "No start or stop codon is given"
XX
PN WO200242322-A2.
XX
PD 30-MAY-2002.
XX
PF 21-NOV-2001; 2001WO-EP013548.
XX
PR 21-NOV-2000; 2000EP-00125524.
XX
PA (LION-) LION BIOSCIENCE AG.
XX
PI Jackson D, Casari G, Suckow J;
XX
DR WPI; 2002-565559/60.
DR P-PSDB; AA015406.
XX
PT Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT useful for inhibiting cellular function of cofactor and for treating
PT metabolic disorders, immunological indications and hormonal dysfunctions.
XX
PS Claim 1; Fig 1; 68pp; English.

Qy　81　GCAGGGTGAAGAAATGAATGAAGAGATTATCCCGTAGAAATTCAGCGAGTCTTTAACAGCCCCTG　140
　　　|||　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　1　GCAGGGTGAAGAAATTAATGAAGAGCTATCCAGTAGAAATTCAGCGAGTTATTTGTCAGCATTT　60

Qy　141　GAGAGCTCCCTGGGGTGCTGGTGACGACAGCACTATCCAGTAGAAGACCTATTCGGCTGGTTTTT　200
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　61　GCGAATTCCATTGATGCTGTGTGATGAGATGCTGAAGAACACATGATGTCTGTTTCTAGAAAT　120

Qy　201　GAGTTGTTGCAGAAGTTGGACCCATTGGAACAACAAGCAAAAGGTGGGATTTAGTTTCTGCATAC　260
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　121　GAGTTGTTGCAGAAGTTGGGACCCACTTGAACACAAGCAAAAGTGGAATTGGGTTTCTGCATAC　180

Qy　261　ACCTTAAATTCAATGTTTTTGGGTTTTATTTGGCAACTCAAGGAGTTAATCCCAAAGAGCAT　320
　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　181　ATATTAAATTCAATGTTTTTGGGCTTATTTGGCAACTCAAGGAGTGAATCCTAAGGAACAT　240

Qy　321　CCAGTGAAGCAGGAACTGGAAAAGAATCAGAGCAGTCTACATGAACAGAGTTAAAGAAAATAACA　380
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　241　TCAGTAAAGCAGGAAATTGGAAGAAATCAGAGTATATATATGAACAGAGTCAAGGAAATAGCA　300

Qy　381　GACAAGAAGAAGAGGCTGCCAAGCTGGACGAGAGGGTGCTGCTTCGAGATTTGTCAAGAAGGCA　440
　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　301　GACAAGAAAAAAGGCTGGCAAGCTGGACAGAGGGTGCAGCTTCAAGAGATTTGTAAGAAATGCC　360

Qy　441　CTCTGGGAACCCAAACACAAAAGACACACCAAAAAGTGCTAATTAAAAAGGGAAAAAGCAAA　497
　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　361　CTCTGGGAACCAAAAACCGAAAAAATGCATCCAAAGTTGCCCTATAAAGGAAAAAAGTAAA　417

RESULT 14
AAL43976/c
ID AAL43976 standard; cDNA; 420 BP.
XX
AC AAL43976;
XX
DT 27-SEP-2002 (first entry)
XX
DE Mammalian nuclear receptor cofactor CF8 reverse complement sequence.
XX
KW Mammalian; gene; 88; nuclear receptor cofactor; CF7; CF8;
KW metabolism regulation; cell homeostasis; cell proliferation;
KW differentiation; pathological cellular aberration;
KW cellular defence mechanism.
XX
OS Mammalia.
XX
PN WO200242322-A2.
XX
PD 30-MAY-2002.
XX
PF 21-NOV-2001; 2001WO-EP013548.
XX
PR 21-NOV-2000; 2000EP-00125524.
XX
PA (LION-) LION BIOSCIENCE AG.
XX

PI   Jackson D, Casari G, Suckow J;
XX
DR   WPI; 2002-566559/60.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.
XX
PS   Claim 1; Fig 2; 68pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of two
CC   mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC   CF8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CF7
CC   and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC   involved in cellular functions, such as: regulation of metabolism and
CC   cell homeostasis, cell proliferation and differentiation, pathological
CC   cellular aberrations, or cellular defence mechanisms. The present cDNA
CC   sequence represents the reverse complement of the mammalian nuclear
CC   receptor cofactor CF8 gene sequence
XX
SQ   Sequence 420 BP; 105 A; 100 C; 62 G; 153 T; 0 U; 0 Other;

Query Match   84.7%; Score 314.6; DB 6; Length 420;
Best Local Similarity   30.3%;
Matches 353; Conservative   0; Mismatches   64; Indels   0; Gaps   0;

Qy   81 GCAGGTGAAGCAAAATGAAATGAAGAATTATCCGGTAGAAAATTCACGAGTCTTAACAGCCCTG 140
                |||||||||||||| |||||||| ||||||||||||| |||||||||||| ||||||||||||
Db   420 GCAGCTGAAGCAAAATTAATGAGGACTATCCAGTAGAAAATTCACGGATTATTATTGTCAGCATTT 361

Qy   141 GAGAGCTCCCTGGGTGCTGTGCTGGACAGACCATGATGAGGGCTGGTTGTTTCTTAGAAAC 200
                |||||| |||||||||||||||||| ||||||||||| ||||||||||| |||| |||||
Db   360 GCGGAATTCCATTGATGCTGTGGATAGAGATGCTGAAGAACATGATGAGTGTCTGTTTCTTAGAAAT 301

Qy   201 GAGTTGTTGCAGGAAGTTGGACCCATTGGAACAAGCAAAGGTGGATTTAGTTTTCTGCATAC 260
                |||||||||||| |||||||||||||||| |||||||||||||||||| |||||||||||
Db   300 GAGTTGTTGCAGAAGTTGGACCCACTGAACAACAGCAAAAGTGGATTTGGTTTTCTGCATAC 241

Qy   261 ACCTTAAATTCAATGTTTGGTTTTATTTGGCAACTCAAGGAGTTAAATCCCAAAGAGCAT 320
                ||||| |||||||||| |||||||| |||||||||||||||||| ||||||||||||||
Db   240 ATATTAAATTCAATGTTTTGGCCTTATTTGGCAACTGAAATCCTAAGGAACAT 181

Qy   321 CCAGTGAAGCAGGAACTCAGAGATCTCTCATGAACAGAGTTAAAGAAAATAACA 380
                |||||||||||||| ||||| |||||||| |||||||| ||||||||||||||
Db   180 TCAGTAGAAGCAGGAAATTGGAAAGAAATCAGAGTTATATAATGGCTGTCAAGGAAATAGCA 121

Qy   381 GACAAGAGAGAGGCTGCCAAGCTGGACAGAGGTGCTGCTTCGAGNTTGTTCAAGAAGGCA 440
                |||||||||||| |||||||||||||||||| ||||||| ||| ||||||||||||||||
Db   120 GACAAGAAAAAGGCTGGCAACGCTGAGAGGTGCAGCTTCAAGATTTGTAAGAAATGCC 61

Qy   441 CTCTCGGGAACCCAAACGAAAAAAGCACACCAAAAAGTGGCTTAATATAAAGGGAAAAGCAAA 497
                |||||||||||||||| ||||||||| ||||||| ||||||||||| |||||||||||| |||
Db   60 CTCTCGGGAACCCAAAACCGAAAAAAATGCATCCAAAAGTTGCCTTATAAAAGGAAAAAAGTAAA 4

PD   06-SEP-2000.
XX
PP   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999; 99US-01224087P.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG01231.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 1235; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors
XX
SQ   Sequence 380 BP; 119 A; 59 C; 95 G; 106 T; 0 U; 1 Other;

Query Match   83.1%; Score 263.2; DB 3; Length 380;
Best Local Similarity   25.3%; Pred. No. 4.8e-51;
Matches 314; Conservative   1; Mismatches   54; Indels   9; Gaps   1;

Qy   15 TGGCGGTCATCATCGTGCCGAACCTATTTCCGGAGACAGGCGTCCACGGTATTGAG------ 68
                |||||||||| || : ||||||||||||||||||||||||||||||||||||||||
Db   3 TGTGCTGGTCGTTGCCGGACCGCTTCCGGAGGAGACTGGGAGTCGGAAGGCGGTGAGGTATTT 62

Qy   69 ---TTGGTCACAATGGCAGGTGAAGAAGAAAATGAAGAAGAGATTAATCCCGTAGAAATTCACGAG 125
                   ||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   63 TTCTAAGGCCATAATAATGGCAGGTGAAGAAAATTAATGAAGAGACTATTCCAGTAGAAATTCACGAG 122

Qy   126 TCTTTAACAGCCCTGGAGAGCTCCCTGGGTGCTGTGGCACACATGCTGAAGACCATGATG 185
                ||||||||||||| |||||| |||||||||||||||| |||||| |||||||||||||||
Db   123 TATTTGTCAGCGTTTGAGAATTCCATTGGTGCTGTGGATGATGCTGAAGAACCATGATG 182

Qy   186 GCTGCTTTCTAGAACCGAGTTGTTGCAGGAGTCGTTGGAACAAGCAAAGGTGGAT 245
                |||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   186 TCTGTTCTTCAGAGAAGTTGTTGCAGGAGTCACTTGAACAAGCAAAAGTGGGAT 245

Qy   246 TTAGTTTTCTGCATACCACCTTAAATTCAATGTTTGGTTTTATTTGGCAACTCAAGGAGTT 305
                |||||| ||||||| |||||||||||||||||| |||||||||||||||||| |||||||
Db   243 TTGGTTTTCTGCATACACATTAAATTCAATGTTTTGGTTATTTGGCAACCCAAGGAGTT 302

Qy   306 AATCCCAAAGAGCATCCAGTGAAGCAGGAACTGAAAAAGAATCAGAGTCTACATGAACAGA 365
                ||||||||||||| ||||||||||||||| |||||||| ||||||| |||| |||||||||
Db   303 AATCCTAAGGAACATCCAGTAAAACAGGAATTGGAAAGAAATCAGAGTATATATGAACAGA 362

Qy   366 GTTAAAGAAAATAACAGAC 383
                ||||||||||||||||||
Db   363 GTCAAGGAAAATAACAGAC 380

RESULT 15
AAC01237
ID   AAC01237 standard; cDNA; 380 BP.
XX
AC   AAC01237;
XX
DT   06-OCT-2000   (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 1235.
XX
OS   Homo sapiens.
KW   gene therapy; chromosome mapping; ss.
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX
PN   EP103401-A2.
XX

Search completed: January 4, 2005, 08:33:22
Job time :579.883 secs

OM protein - protein search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.   GenCore version 5.1.6

Run on: January 4, 2005, 08:33:30

Sequence: 1 MAGEEMNEDYPVEIHESLTA...........LWEPKRKSTPKVANKGKSKH 141
Title: US-09-701-618A-4
Perfect Score: 715
Title:

Scoring table: BLOSUM62
                   Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   UniProt_02:*
                   1: uniprot_sprot:*
                   2: uniprot_trembl:*

                   (without alignments)
                   ; Search time 63.5 Seconds
                   1277.604 Million cell updates/sec

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	715	100.0	141	2	Q61368	mus musculu
2	704	98.5	141	2	O35473	m small uni
3	684	95.7	141	2	Q7TSU0	cricetulus
4	646	90.3	141	2	Q13901	homo sapien
5	515	72.0	111	2	Q7L8E1	homo sapien
6	182.5	25.5	141	2	Q86F82	schistosoma
7	154.5	21.6	217	2	Q93VA9	arabidopsis
8	144	20.1	204	2	Q6BJZ2	debaryomyce
9	142	19.9	133	2	O74469	schizosacch
10	137	19.2	207	2	Q6YX09	oryza sativ
11	137	19.2	207	2	BAC99871	oryza sat
12	134	18.7	450	2	Q7SER2	neurospora
13	131.5	18.4	169	2	Q70IZ7	anopheles g
14	129	18.0	194	2	Q6C2H9	yarrowia li
15	128	17.9	159	2	Q9VXL4	drosophila
16	124	17.3	149	2	Q8IL18	plasmodium
17	121.5	17.0	133	2	Q9N3J4	caenorhabdi
18	100	14.0	184	2	P3B801	saccharomyc
19	98	13.7	149	2	Q6FUI1	candida gla
20	97.5	13.6	654	2	Q8L3P4	arabidopsis
21	90.5	12.7	244	2	Q9HGN2	schizosacch
22	90	12.6	395	2	Q6CQF7	kluyveromyc
23	89.5	12.5	364	2	Q86RM8	plasmodium
24	88	12.3	830	2	Q872F6	neurospora
25	87.5	12.2	177	2	Q6CMX3	kluyveromyc
26	87.5	12.2	229	2	Q80DB1	oryza sativ
27	87.5	12.2	479	2	Q6FI82	oryza sativ
28	87.5	12.2	479	2	Q9NQ22	homo sapien
29	86.5	12.1	508	1	P75064	mycoplasma
30	86.5	12.1	595	2	Q86DH4	plasmodium
31	86.5	12.1	621	2	Q86RM2	plasmodium
32	86.5	12.1	625	2	Q86RN6	plasmodium
33	86.5	12.1	639	2	Q86RM4	plasmodium
34	86.5	12.1	732	2	Q86RM7	plasmodium
35	86.5	12.1	849	2	Q86R71	plasmodium
36	86.5	12.1	849	2	Q86RM3	plasmodium
37	86.5	12.1	855	2	Q86DH2	plasmodium
38	86.5	12.1	856	2	Q86RM6	plasmodium
39	86	12.0	915	2	O32863	mycoplasma
40	86	12.0	1208	2	Q9VXU2	drosophila
41	85.5	12.0	1398	2	Q835M0	enterococcu
42	85.5	12.0	328	2	Q9VXU1	drosophila
43	85.5	12.0	370	2	Q6MRI4	bdellovibri
44	85.5	12.0	370	2	Cae77774	bdellovib
45	85	11.9	1997	2	Q8IAP3	plasmodium

ALIGNMENTS

RESULT 1
ID Q61368   PRELIMINARY;   PRT;   141 AA.
AC Q61368;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE C1D protein.
GN Name=C1d; Synonyms=C1D;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI; TISSUE=Ascites tumor;
RX MEDLINE=98136194; PubMed=9469821;
RA Nehls P., Kecks T., Greferath R., Spiess E., Glaser T., Rothbarth K.,
RA Stammer H., Werner D.;
RT "cDNA cloning, recombinant expression and characterization of
RT polypeptides with exceptional DNA affinity.";
RL Nucleic Acids Res. 26:1160-1166(1998).
DR EMBL; X95591; CAA64844.1; -.
DR MGD; MGI:1927354; C1d.
DR GO; GO:0017053; C:transcriptional repressor complex; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0016922; F:ligand-dependent nuclear receptor interacto...; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0003714; F:transcription corepressor activity; IDA.
DR GO; GO:0016481; P:negative regulation of transcription; IDA.
DR InterPro; IPR011082; C1D.
DR Pfam; PF07493; C1D; 1.
SQ SEQUENCE 141 AA; 16045 MW; C9AF68033A097161 CRC64;

Query Match   100.0%; Score 715; DB 2; Length 141;
Best Local Similarity 100.0%;   Pred. No. 1.3e-55;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK 120

Qy 121 ALWEPKRKSTPKVANKGKSKH 141
         |||||||||||||||||||||
Db 121 ALWEPKRKSTPKVANKGKSKH 141

RESULT 2
ID O35473   PRELIMINARY;   PRT;   141 AA.
AC O35473;

AC   O35473;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Small unique nuclear receptor co-repressor (Nuclear DNA-binding
DE   protein) (Mus musculus 10 days neonate skin cDNA, RIKEN full-length
DE   enriched library, clone:4732439J08 product:nuclear DNA binding
DE   protein, full insert sequence) (Mus musculus 12 days embryo embryonic
DE   body between diaphragm region and neck cDNA, RIKEN full-length
DE   enriched library, clone:9430096A21 product:nuclear DNA binding
DE   protein, full insert sequence)
GN   Name=C1d; Synonyms=SUN-CoR;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98070763; PubMed=9405624;
RA   Zamir I., Dawson J., Lavinsky R.M., Glass C.K., Rosenfeld M.G.,
RA   Lazar M.A.;
RT   "Cloning and characterization of a corepressor and potential component
RT   of the nuclear hormone receptor repression complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:14400-14405(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CZECH II;
RC   TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RX   MEDLINE=22382257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CZECH II;
RC   TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and skin;
RX   MEDLINE=99279253; PubMed=10349636;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and skin;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [6]
RP   SEQUENCE FROM N.A.

RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and skin;
RA   The FANTOM Consortium;
RA   The RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and skin;
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and skin;
RX   MEDLINE=20530913; PubMed=11076861;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka K., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RC   TISSUE=Embryonic body between diaphragm region and neck, and skin;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto N., Hiraoka Y., Hirozane T.,
RA   Hori F., Inotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kirihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai T., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi M., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF031426; AAC53520.1; -.
DR   EMBL; BC005436; AAH05436.1; -.
DR   EMBL; AK028702; BAC26075.1; -.
DR   EMBL; AK035169; BAC28967.1; -.
DR   MGD; MGI:1927354; C1d.
DR   GO; GO:0017053; C:transcriptional repressor complex; IDA.
DR   GO; GO:0003677; F:DNA binding; IDA.
DR   GO; GO:0016922; F:ligand-dependent nuclear receptor interacto...; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0003714; F:transcription corepressor activity; IDA.
DR   GO; GO:0016481; P:negative regulation of transcription; IDA.
DR   InterPro; IPR011082; C1d.
DR   Pfam; PF07493; C1d; 1.
KW   DNA-binding; Receptor.
SQ   SEQUENCE 141 AA; 15945 MW; C9AF68033A0FD04E CRC64;

Query Match 98.5%; Score 704; DB 2; Length 141;
Best Local Similarity 98.6%;
Matches 139; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy   1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMAVSRNELLQKLDPLEQAKVDLVSA 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMAVSRNELLQKLDPLEQAKVDLVSA 60

Qy  61 YTLNSMFWVYLLATQGVNPKEHPVKQELERIRVYMRVKEITDKKAAKLDRGAASRFVKK 120
```

```
Db
Qy    61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKN 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ALWEPKRKSTPKVANKGKSH 141

Qy   121 ALWEPKRKSTPKVANKGKSH 141
          ||||||||:|:||||||||
Db   121 ALWEPKAKSTPKVANKGKSH 141
```

RESULT 3
Q7TSU0

```
ID   Q7TSU0;          PRELIMINARY;      PRT;    141 AA.
AC   Q7TSU0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Nuclear DNA-binding protein C1D.
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Guang L., Maabumi S., Maru Y.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY302220; AAP43113.1; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   DNA-binding.
SQ   SEQUENCE   141 AA;  15945 MW;  0CFB6477685566 CRC64;

Query Match             95.7%;  Score 684;  DB 2;  Length 141;
Best Local Similarity   95.0%;              Pred. No. 7.3e-53;
Matches  134;  Conservative   4;  Mismatches   3;  Indels   0;  Gaps   0;

Qy    1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60
         |||:||||| :||||||| ||||||||||||||||:|||||||||||||||||||||||
Db    1 MAGGEMNEDVPVEIHESLSALESSLGAVDDMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK 120

Qy  121 ALWEPKRKSTPKVANKGKSH 141
         ||||||:|:||:||||||||
Db  121 ALWEPKQKNTPNVANKGKSH 141
```

RESULT 4
Q13901

```
ID   Q13901;          PRELIMINARY;      PRT;    141 AA.
AC   Q13901;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   C1D protein (Nuclear DNA-binding protein).
GN   Name=C1D;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Term placenta;
RX   MEDLINE=98136194; PubMed=9469821;
RA   Nehls P., Keeks T., Greferath R., Spiess E., Glaser T., Rothbarth K.,
RA   Stamner H., Werner D.;
RT   "cDNA cloning, recombinant expression and characterization of
RT   polypeptides with exceptional DNA affinity.";
RL   Nucleic Acids Res. 26:1160-1166(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow, Lung, and Uterus;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X95592; CAA64845.1; -.
DR   EMBL; BC009584; AAH09584.1; -.
DR   EMBL; BC009584; AAH09584.1; -.
DR   EMBL; BC009589; AAH09589.1; -.
DR   EMBL; BC009589; AAH09589.1; -.
DR   EMBL; BC016284; AAH16284.1; -.
DR   EMBL; BC016284; AAH16284.1; -.
DR   EMBL; BC005235; AAH05235.1; -.
DR   EMBL; BC005235; AAH05235.1; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0003677; F:DNA binding; TAS.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   DNA-binding.
SQ   SEQUENCE   141 AA;  16019 MW;  9976A3BBD5620D63 CRC64;

Query Match             90.3%;  Score 646;  DB 2;  Length 141;
Best Local Similarity   90.0%;              Pred. No. 1.7e-49;
Matches  126;  Conservative   7;  Mismatches   7;  Indels   0;  Gaps   0;

Qy    1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60
         ||||:|||||||||||:||:|:||:|||||:||||||:||||||||||||||||||||||
Db    1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKN 120
         ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAGKLDRGAASRFVKN 120

Qy  121 ALWEPKRKSTPKVANKGKSH 140
         ||||||:|:|:||||||||
Db  121 ALWEPKSKNASKVANKGKSH 140
```

RESULT 5
Q7L8E1

```
ID   Q7L8E1;          PRELIMINARY;      PRT;    111 AA.
AC   Q7L8E1;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
```

```
DE   BA19501.1 (Novel protein similar to C1D DNA-binding protein).
GN   Name=bA19501.1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lawlor S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL354750; CAD13204.1; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   DNA-binding;
SQ   SEQUENCE   111 AA;  12646 MW;  F563EA631CEC6EE1 CRC64;

Query Match             92.7%;  Score 515;  DB 2;  Length 111;
Best Local Similarity
Matches  102;  Conservative   2;  Mismatches   6;  Indels   0;  Gaps   0;

Qy  31 MLKTMMAVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERI 90
        |||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MLKTTMSVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELERI 60

Qy  91 RVYMNRVKEITDKKKAAKLDRGAASRFVKKALWEPKRKSTPKVANKGKSK 140
        ||||||||||||:||:|||||||||||||:|||||||:|||:|||||||
Db  61 RVYMNRVKEITDDKKAGKLDRGAASRFVKNALWEPKPKNASKVANKGKSK 110


RESULT 6
ID   Q86F82             PRELIMINARY;      PRT;   141 AA.
AC   Q86F82;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Clone ZZD1007 mRNA sequence.
OS   Schistosoma japonicum (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC   Schistosomatoidea; Schistosomatidae; Schistosoma.
OX   NCBI_TaxID=6182;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22879925; PubMed=12973349;
RA   Hu W., Yan Q., Shen D.K., Liu F., Zhu Z.D., Song H.D., Xu X.R.,
RA   Wang Z.J., Rong Y.P., Zeng L.C., Wu J., Zhang X., Wang J.J., Xu X.N.,
RA   Wang S.Y., Fu G., Zhang X.L., Wang Z.Q., Brindley P.J., McManus D.P.,
RA   Xue C.L., Feng Z., Chen Z., Han Z.G.;
RT   "Evolutionary and biomedical implications of a Schistosoma japonicum
RT   complementary DNA resource.";
RL   Nat. Genet. 35:139-147(2003).
DR   EMBL; AY222982; AAP06005.1; -.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
SQ   SEQUENCE   141 AA;  15966 MW;  DBEFB20FA1894BA0 CRC64;

Query Match             28.4%;  Score 182.5;  DB 2;  Length 141;
Best Local Similarity
Matches   38;  Conservative  33;  Mismatches  60;  Indels   3;  Gaps   2;

Qy   1 MAGEMENEDYPVEIHESLLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60
        ::|:| || :::||::.| |:|::||  ||||||||||||||||||||||||||||||
Db   1 MGDSSIFDEIPKEISSQLVSFSEATDDVEQLVNKISSPSNNS-SNEVSGLDTVKSELSLC 59

Qy  61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAK---LDRGAASRFV 118
        |  |::|  || : : |  |: |:|:|:|  ::: :| : : | ||   |||||||||
Db  60 YAFNALFMYLRCNGVETQSHPIMQELDRVMNALKRCRSLVEREVSARLTLDKEATTRFV 119

Qy 119 KKALWEPKRKSTPK 132
        | ||| :   | |
Db 120 KHALWKSAHTKTK 133


RESULT 7
ID   Q93VA9             PRELIMINARY;      PRT;   217 AA.
AC   Q93VA9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   AT5g25080/T11H3_90 (Hypothetical protein).
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim C.J., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA   Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA   Satou M., Seki M., Toriumi M., Sakurai T.,
RA   Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA   Ecker J.R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22088475; PubMed=12093376;
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 3:RESEARCH0029-RESEARCH0029(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Chen H., Kim C.J., Koesema E., Meyers M.C.,
RA   Bahn J., Bowser L., Carninci P., Chung M.K., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA   Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T.,
RA   Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA   Yu G., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY050469; AAK91482.1; -.
DR   EMBL; AY087408; AAM64957.1; -.
DR   EMBL; AF378870; AAK55673.1; -.
DR   InterPro; IPR011082; C1D.
DR   Pfam; PF07493; C1D; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   217 AA;  23748 MW;  DB44F167BD773DE4 CRC64;

Query Match             30.6%;  Score 154.5;  DB 2;  Length 217;
Best Local Similarity                         Pred. No. 1.2e-05;
Matches   41;  Conservative  30;  Mismatches  46;  Indels  17;  Gaps   3;

Qy  22 ESSLGAVDD----------MLKTMMAVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYL 71
        |::| ||||          ||: | | :|: :  ::|||||  :| || : :::| ||
Db   1 ESAIEAVNQTLAYLKELKPLEQMLTLAEPEVLAAMQPLQRAKTMHLLAERATTLYELRL 71

Qy  72 ATQGVNPKEHPVKQELERIRVYMNRVKEITDKKK-------AAKLDRGAASRFVKKALWEPK 126
        .| ||:|::| ||:|:|| :|| |. ::  : :       ||||| |:  ||: |::
Db  72 RCTGVDPDDHRVKSEIRINVYREKFQKCVDDQSKGPLRPTTVLNRQAATRFIEHSL--PD 135

Qy 127 RKSTPKVANKGKSK 140
        : ||:|:  | |
Db 136 LTSTQKQSIRDLSK 149
```

RESULT 8

ID   Q6BJZ2   PRELIMINARY;   PRT;   204 AA.
AC   Q6BJZ2;
DT   01-OCT-2004 (TrEMBLrel. 28, Created)
DT   01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to CA0533I1PF19872.
DE   ORFNames=DEHA0F27797g;
GN   ORFNames=DEHA0F27797g;
OS   Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX   NCBI_TaxID=4959;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RG   GENOLEVURES;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
SQ   SEQUENCE   204 AA;   22569 MW; 724A599FC34D0C9C CRC64;

Query Match   20.1%; Score 144; DB 2; Length 204;
Best Local Similarity   32.7%; Pred. No. 9.5e-05;
Matches   35; Conservative 26; Mismatches 42; Indels 4; Gaps 2;

Qy   20 ALESSLGAVDDMLKTMMAVSRNELLQKLD-PLEQAKVDLVSAYTLNSMFWVYLLATQGVNP 78
            :| ||: | |:| ||| |:|||:|||| ||||||||| || ||||| | : | ||
Db   13 SLDNSVDQLTEKLEPILKKSLEEKIAASDSQVERIKIVNNYSVVLISILFSYLKTLGINT 72

Qy   79 KEHPVKQELERIRVYMNRVKEITDK----KKAAKLDRGAASRFVKKAL 122
            |:|: ||:||||||||||: | || : |||||||||||||||||||
Db   73 DQHPIMKELTRIKLYMKRYKELEAKLIASKDTSKEDAEAARTFIQNTL 119

RESULT 9

ID   O74469   PRELIMINARY;   PRT;   133 AA.
AC   O74469;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SPCC1739.07 protein.
GN   Name=SPCC1739.07;
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972h-;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schafer M., Muller-Auer S.,
RA   Gabel C., Fuchs M., Dusterhoft A., Fritzc C., Holzer E., Moestl D.,
RA   Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
RA   Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
DR   EMBL; AL031540; CAA22078.1; -.
DR   PIR; T41115; T41115.
DR   GeneDB_SPombe; SPCC1739.07; -.
DR   InterPro; IPR011082; CID.
DR   Pfam; PF07493; CID; 1.
SQ   SEQUENCE   133 AA;   15068 MW; DD33501508D60F1 CRC64;

Query Match   19.9%; Score 142; DB 2; Length 133;
Best Local Similarity   28.6%; Pred. No. 8.7e-05;
Matches   38; Conservative 24; Mismatches 57; Indels 14; Gaps 3;

Qy   13 EIHESLTALESSLGAVDDMLKTMM-AVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYL 71
            : |: | |||||||||||| | | |||||| | ||||||||| || |||||| ||
Db   4 EYSELFERLNKQLDNVEDVLKPILKDAESIFELAEGKSELEQAKLYITMSVAINSTLYSFY 63

Qy   72 ATQGVNPKEHPVKQELERIRVYMNRV----KEITDKKKAAKLDRGAASRFVKKALMEPKR 127
            | :| |||||||||||||||||||| |:| | ||||||||||||||||| ||
Db   64 KLNGIDASERPVMQELQRVKNYISKIQQAEKNVNPKTEAVNTSNAAIS------SSS 114

RESULT 10

ID   Q6YX09   PRELIMINARY;   PRT;   207 AA.
AC   Q6YX09;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein OSJNBa0091D16.5.
GN   Name=OSJNBa0091D16.5;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sasaki T., Matsumoto T., Katayose Y.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005731; BAC99871.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   207 AA;   22203 MW; 0B0161AD8190E071 CRC64;

Query Match   19.2%; Score 137; DB 2; Length 207;
Best Local Similarity   32.2%; Pred. No. 0.0004;

Qy   128 KSTPKVANKGKSK 140
            :||||
Db   115 SNRPKVAKDAATR 127

```
Matches  37;  Conservative  24;  Mismatches  44;  Indels  10;  Gaps  3;

Qy   18 LTALESSLGA---VDDMLKTMMAVSRN--ELLQKLDLPLEQAKVDLVSAYTLNSMFWVYLA 72
Db   18 VSAAEDTLAAAESVGDHLAEMVAAAGEDPDAIAELPLRRARAFLAMAKAAASLFAVRLR 77

Qy   73 TQGVNPKEHPVKQELERIRVYMNRVKEITDKKKA------AKLDRGAASRFVKKAL 122
Db   78 CSGVDPDEHPIKKEFERLSIWEEKLNRFEDWDKAPLRPTTVNTQAAARFIGHSL 132
```

RESULT 11
BAC99871

```
ID   BAC99871    PRELIMINARY;    PRT;   207 AA.
AC   BAC99871;
DT   02-MAR-2004  (TrEMBLrel. 27, Created)
DT   02-MAR-2004  (TrEMBLrel. 27, Last sequence update)
DT   02-MAR-2004  (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein OSJNBa0091D16.5.
GN   OSJNBa0091D16.5.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza; Oryza sativa.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sasaki T., Matsumoto T., Katayose Y.;
RT   "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
RT   clone:OSJNBa0091D16.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005731; BAC99871.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE  207 AA;  22203 MW;  0B0161AD8190B071 CRC64;
```

Query Match 19.2%; Score 137; DB 2; Length 207;
Best Local Similarity 32.2%; Pred. No. 0.0004;

RESULT 12
Q7SER2

```
ID   Q7SER2    PRELIMINARY;    PRT;   450 AA.
AC   Q7SER2;
DT   01-MAR-2004  (TrEMBLrel. 26, Created)
DT   01-MAR-2004  (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004  (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
DE   Name=NCU02154.1;
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA   Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA   Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA   Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA   Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA   Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA   Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA   Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA   Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA   Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A.,
RA   DeSouza C.C., Glass L., Orbach M.J., Berglund J.A., Voelker R.,
RA   Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA   Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA   Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
RT   "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL   Nature 0:0-0(2003).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABX01000029; EAA35277.1; -.
DR   InterPro; IPR010625; CHCH.
DR   Pfam; PF06747; CHCH; 1.
DR   Pfam; PF07493; CID; 1.
KW   Hypothetical protein.
SQ   SEQUENCE  450 AA;  48201 MW;  C0C8F30E5AD0810 CRC64;
```

Query Match 18.7%; Score 134; DB 2; Length 450;
Best Local Similarity 29.2%; Pred. No. 0.0018;
Matches 40; Conservative 35; Mismatches 54; Indels 8; Gaps 5;

```
Qy   9 DYPVEIHESLITALESSLGAVDDMLKTMMAVSRNELLQKLDLPLEQAKVDLVSAYTLNSMFW 68
Db   2 DVP-DITPSLEKLNVDLDQLEAALKPVLG-DVGDVSSKLPLLDKAKLYVMVTYAIESILF 59

Qy   69 VYLATQGVNPKEHPVKQELERIRVYMNRVKEITD--KKKAAKLDRGAASRFVKKALWEP- 125
Db   60 SSLRLNGVDAKEHAIFTELTRVRQYFEKIQNIENPQEERQCTVNKEAAARFIRSDLADDE 119

Qy   126 --KRKSTPKVANKGKSK 140
Db   120 AIKQKLTELIA-KEKAK 135
```

RESULT 13
Q7Q127

```
ID   Q7Q127    PRELIMINARY;    PRT;   169 AA.
AC   Q7Q127;
DT   01-MAR-2004  (TrEMBLrel. 26, Created)
DT   01-MAR-2004  (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004  (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=agCG52216; ORFNames=ENSANGG00000016976;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008807; EAA04511.1; -.
NON_TER   1   1
SQ   SEQUENCE  169 AA;  19590 MW;  69C2764F057A93C2 CRC64;
```

Query Match 18.4%; Score 131.5; DB 2; Length 169;
Best Local Similarity 28.9%; Pred. No. 0.0098;
Matches 37; Conservative 27; Mismatches 55; Indels 9; Gaps 4;

```
Qy   22 ESSLGAVDDMLKTMMAVSRN-ELLQKLDLPLEQAKVDLVSAYTLNSMFWVYLATQG-VNPK 79
Db   37 ENLMQAIDQIQVTLANVREYEKYPNFFAEEKVRFDTLCAFCVNSLFWMHEKMLGRPNTV 96

Qy   80 EHPVKQELERIRVYMNRVKEITDK-KKAAKLDRGAASRFVKKALWE------PKRKSTPK 132
Db   97 MDDIKADLDRVREAMKRLQTIHDNLIKRPRLDPAAQRFVRAGLIYDANNAESDKQNDAPP 156
```

Qy 133 VANKGKSK 140
         |:|
Db 157 NKKKRREK 164

RESULT 14

Qy   Q6C2H9
Db   Q6C2H9

ID   Q6C2H9;   PRELIMINARY;   PRT;   194 AA.
AC   Q6C2H9;
DT   01-OCT-2004 (TrEMBLrel. 28, Created)
DT   01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to DEHA0P27797g Debaryomyces hansenii.
GN   ORFNames=YALI0F07755g;
OS   Yarrowia lipolytica (Candida lipolytica).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=4952;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RG   GENOLEVURES;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.-F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
DR   EMBL; CR382112; CAG77940.1; -.
SQ   SEQUENCE   194 AA;   21520 MW;   D67D7D9F8068BB3B CRC64;

Query Match   18.0%; Score 129; DB 2; Length 194;
Best Local Similarity   28.3%; Pred. No. 0.0019;
Matches   32; Conservative   22; Mismatches 35; Indels 24; Gaps   3;

Qy   30 DMLKTMMAVSRN-------ELLQ-------KLDPLEQAKVDLVSAYTLNSMFWVYLA 72
               |:|::||::|| |||| |:||| :||||| ||:||| :::|:|||:
Db   3 DQIEDVLELSHNLQDVTSELSQIKMIDFKGVAQLPPLEQAQFYSKLAYVTNSAMFAFIL 62

Qy   73 TQGVNPKEHPVKQELERIRVYMNRVKEITDK------------KKAAKLDRGAASRFV 118
               |::::|:|:|||:|:: ||:|:|:|:|||:||| |||||||||||:|
Db   63 ASGDPKTHPIMKLDRVKTYMGKVAHAEGKPGPARKDERNTKVDPAAKRII 115

RESULT 15

Qy   Q9VXL4
Db   Q9VXL4

ID   Q9VXL4;   PRELIMINARY;   PRT;   159 AA.
AC   Q9VXL4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG8928-PA.
GN   ORFNames=CG8928;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;

RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIB99;
RG   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a

```
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FLYBASE;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FLYBASE;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB003500; AAF48545.1; -.
DR   IntAct; Q9VXL4; -.
DR   FlyBase; FBgn0030711; CG8928.
SQ   SEQUENCE 159 AA; 18138 MW; 38CB9E302A75BFD6 CRC64;

Query Match             17.9%;  Score 128;  DB 2;  Length 159;
Best Local Similarity   29.2%;  Pred. No. 0.0019;
Matches   42;  Conservative  21;  Mismatches  47;  Indels  34;  Gaps   4;

Qy   21 LESSLGAVDDM---LKTMMAVSRNELLQ-----------------KLDPLEQAKVDLVSAYTLN 64
         |::||       |||                           :  | |::|
Db   18 LDTSLREDENMQHILKTF--YSSIELLEADTEKALLALQAERTLNTNEQIKLDSYLVYLNS 75

Qy   65 SMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITTDKKKAAKLDRGAASRF------ 117
         ::||:||:|    |  |       : |:|      |::|      |||
Db   76 TLFFIYLKLQGEDASNHAVMHDLRRTRDLLARDKKINDALAAPRLDMPAAKRFIAAGTHT 135

Qy  118 --------VKKALWEPKRKSTPK 132
                 : :          |||
Db  136 RFVDMNGVMVSEKQYNKSKQETPK 159

Search completed: January 4, 2005, 13:44:43
Job time : 67.5 secs
```

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: January 4, 2005, 08:13:30 ; Search time 66 Seconds
(without alignments)
766.376 Million cell updates/sec

Title: US-09-701-618A-4
Perfect score: 715
Sequence: 1 MAGEEMNEDYPVEIHESLTA.........LWEPKRKSTPKVANKGKSKH 141

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_23Sep04:*
                1: geneseqp19980s:*
                2: geneseqp19990s:*
                3: geneseqp2000s:*
                4: geneseqp2001s:*
                5: geneseqp2002s:*
                6: geneseqp2003as:*
                7: geneseqp2003bs:*
                8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	715	100.0	141	3	AAY51025	Aay51025 Murine C1
2	704	98.5	141	7	ADE59762	Ade59762 Rat Prote
3	646	90.3	141	3	AAY51024	Aay51024 Human C1D
4	646	90.3	141	7	ADE59764	Ade59764 Human Pro
5	615	86.0	140	5	ADI28083	Adi28083 Human nuc
6	615	86.0	141	5	AAG49052	Aag49052 Arabidops
7	579	81.0	140	5	AAO15406	Aao15406 Mammalian
8	485	67.8	102	3	AAO15405	Aao15405 Mammalian
9	235	32.9	54	8	AAG01231	Aag01231 Human sec
10	154.5	21.6	217	3	AAO54730	Abo54730 Human gen
11	154.5	21.6	217	3	AAG49051	Aag49051 Arabidops
12	154.5	21.6	217	3	AAG24127	Aag24127 Arabidops
13	154.5	21.6	253	3	AAG49050	Aag49050 Arabidops
14	146.5	20.5	256	3	AAG24126	Aag24126 Arabidops
15	146.5	20.5	177	3	AAG24128	Aag24128 Arabidops
16	128	17.9	159	4	AAG24128	Aag24128 Arabidops
17	112.5	15.7	232	8	AAB64082	Aag49052 Arabidops
18	100	14.0	184	7	ADP99133	Aab64082 Drosophil
19	100	14.0	184	7	ABR53602	Adp99133 C. albica
20	87.5	12.2	479	3	ADK64282	Abr53602 Protein s
21	87.5	12.2	479	3	AAY92242	Adk64282 Disease t
22	87.5	12.2	479	5	ABB05690	Aay92242 Human can
23	87.5	12.2	479	5	ABB83498	Abb05690 Human nuc
24	87.5	12.2	479	7	ABP64721	Abb83498 Large hum
25	86.5	12.1	508	6	ABU36173	Abp64721 Human pro
					ADJ69180	Adj69180 Human hea
					ABU36173	Abu36173 Protein e
26	86	12.0	1294	4	ABB63502	Abb63502 Drosophil
27	85.5	12.0	295	7	ADH85852	Adh85852 Enterococ
28	83.5	11.7	1055	2	AAW98865	Aaw98865 H. pylori
29	82.5	11.5	2285	2	AAW98149	Aaw98149 Bacillus
30	82	11.5	307	7	ADC97413	Adc97413 B. faeciu
31	81	11.3	131	5	ABP02625	Abp02625 Human ORF
32	81	11.3	345	3	AAG23880	Aag23880 Arabidops
33	80.5	11.3	570	2	AAW57879	Aaw57879 C. felis
34	80.5	11.3	570	2	AAW57860	Aaw57860 C. felis
35	80.5	11.3	570	4	AAE12915	Aae12915 Ctenoceph
36	80.5	11.3	595	2	AAW57859	Aaw57859 C. felis
37	80.5	11.3	595	2	AAW57861	Aaw57861 C. felis
38	80.5	11.3	595	4	AAE12914	Aae12914 Ctenoceph
39	80.5	11.3	1094	5	ABP73717	Abp73717 Candida a
40	80.5	11.3	1976	7	ADE63514	Ade63514 Rat Prote
41	80.5	11.3	1976	7	ADE63518	Ade63518 Rat Prote
42	80	11.2	1027	2	AAR43699	Aar43699 Protein L
43	80	11.2	1027	2	AAR42203	Aar42203 Protein L
44	79	11.0	645	2	AAY35355	Aay35355 Amino aci
45	78.5	11.0	570	2	AAW57878	Aaw57878 C. felis

ALIGNMENTS

RESULT 1
AAY51025

ID   AAY51025 standard; protein; 141 AA.
XX
AC   AAY51025;
XX
DT   17-MAR-2000 (first entry)
XX
DE   Murine C1D protein.
XX
KW   C1D; murine; apoptosis; tumour; gene therapy; treatment.
XX
OS   Mus sp.
XX
XX
PN   DE19824811-A1.
XX
PD   09-DEC-1999.
XX
XX
PF   03-JUN-1998; 98DE-01024811.
XX
PR   03-JUN-1998; 98DE-01024811.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI   Rothbarth K, Stammer H, Werner D;
XX
DR   WPI; 2000-063506/06.
DR   N-PSDB; AA243928.
XX
PT   Inducing apoptosis by overexpressing the C1D gene, particularly for
PT   treating tumors.
XX
PS   Claim 3; Fig 2; 10pp; German.
XX
CC   This invention describes a novel method for inducing apoptosis which
CC   comprises overexpressing the C1D gene (I). The method is particularly
CC   used to treat tumors and can also be used in gene therapy. The method has
CC   no side effects on normal cells (contrast known methods of inducing
CC   apoptosis such as cytotoxins and radiation), and may be effective on
CC   cells resistant to conventional treatments. Overexpression of (I) is
CC   sufficient itself to induce apoptosis but the effect may be increased
CC   when used in combination with other anti-tumor methods. When cells
CC   transfected with (I) undergo apoptosis, they release factors that kill
CC   neighboring, non-transfected cells (bystander effect). This sequence
CC   represents the murine C1D protein described in the method of the
CC   invention
XX
SQ   Sequence 141 AA;

RESULT 2

```
Query Match     100.0%;  Score 715;  DB 3;  Length 141;
Best Local Similarity  100.0%;  Pred. No. 4e-66;
Matches 141;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1  MAGEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  MAGEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA   60

Qy  61  YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61  YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK  120

Qy 121  ALWEPKRKSTPKVANKGKSKH  141
         |||||||||||||||||||||
Db 121  ALWEPKRKSTPKVANKGKSKH  141
```

```
RESULT 2
ID  ADE59762  standard; protein; 141 AA.
XX
AC  ADE59762;
XX
DT  29-JAN-2004  (first entry)
XX
DE  Rat Protein AAH05436, SEQ ID NO 5658.
XX
KW  Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
    chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS  Rattus norvegicus.
XX
PN  WO200301647S-A2.
XX
PD  27-FEB-2003.
XX
PF  14-AUG-2002;  2002WO-US025765.
XX
PR  14-AUG-2001;  2001US-0312147P.
PR  01-NOV-2001;  2001US-0346382P.
PR  26-NOV-2001;  2001US-0333347P.
XX
PA  (GEHO )  GEN HOSPITAL CORP.
PA  (FARB )  BAYER AG.
XX
PI  Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR  WPI; 2003-268312/26.
DR  GENBANK; AAH05436.
XX
PT  New composition comprising two or more isolated polypeptides, useful for
PT  preparing a medicament for treating pain in an animal.
XX
PS  Claim 1; Page; 1017pp; English.
XX
CC  The invention discloses a composition comprising two or more isolated rat
CC  or human polynucleotides or a polynucleotide which represents a fragment,
CC  derivative or allelic variation of the nucleic acid sequence. Also
CC  claimed are a vector comprising the novel polynucleotide, a host cell
CC  comprising the vector, a method for identifying a nucleotide sequence
CC  which is differentially regulated in an animal subjected to pain and a
CC  kit to perform the method, an array, a method for identifying an agent
CC  that increases or decreases the expression of the polynucleotide sequence
CC  that is differentially expressed in neuronal tissue of a first animal
CC  subjected to pain, a method for identifying a compound which regulates
CC  the expression of a polynucleotide sequence which is differentially
CC  expressed in an animal subjected to pain, a method for identifying a
CC  compound that regulates the activity of one or more of the novel
CC  polynucleotides, a method for producing a pharmaceutical composition, a
CC  method for identifying a compound or small molecule that regulates the
CC  activity in an animal of one or more of the polypeptides given in the
CC  specification, a method for identifying a compound useful in treating
CC  pain and a pharmaceutical composition comprising the one or more
CC  polypeptides or their antibodies. The polynucleotide or the compound that
CC  modulates its activity is useful for preparing a medicament for treating
CC  pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC  injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC  therapy). The sequence presented is a rat protein (shown in Table 2 of
CC  the specification) which is differentially expressed during pain. Note:
CC  The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic form directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 141 AA;
```

RESULT 3

```
Query Match      98.5%;  Score 704;  DB 7;  Length 141;
Best Local Similarity  98.6%;  Pred. No. 5.5e-65;
Matches 139;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy   1  MAGEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  MAGEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA   60

Qy  61  YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK  120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61  YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKN  120

Qy 121  ALWEPKRKSTPKVANKGKSKH  141
         ||||| |||||||||||||||
Db 121  ALWEPKTAKSTPKVANKGKSKH  141
```

```
RESULT 3
ID  AAY51024  standard; protein; 141 AA.
XX
AC  AAY51024;
XX
DT  17-MAR-2000  (first entry)
XX
DE  Human C1D protein.
XX
KW  C1D; human; apoptosis; tumour; gene therapy; treatment.
XX
OS  Homo sapiens.
XX
PN  DE19824811-A1.
XX
PD  09-DEC-1999.
XX
PF  03-JUN-1998;  98DE-01024811.
XX
PR  03-JUN-1998;  98DE-01024811.
XX
PA  (DEKR-)  DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI  Rothbarth K,  Stammer H,  Werner D;
XX
DR  WPI; 2000-063506/06.
DR  N-PSDB; AAZ43927.
XX
PT  Inducing apoptosis by overexpressing the C1D gene, particularly for
PT  treating tumors.
XX
PS  Claim 3; Fig 1; 10pp; German.
XX
CC  This invention describes a novel method for inducing apoptosis which
CC  comprises overexpressing the C1D gene (I). The method is particularly
CC  used to treat tumors and can also be used in gene therapy. The method has
CC  no side effects on normal cells (contrast known methods of inducing
CC  apoptosis such as cytotoxins and radiation), and may be effective on
CC  cells resistant to conventional treatments. Overexpression of (I) is
CC  sufficient itself to induce apoptosis but the effect may be increased
CC  when used in combination with other anti-tumor methods. When cells
CC  transfected with (I) undergo apoptosis, they release factors that kill
```

CC   neighboring, non-transfected cells (bystander effect). This
CC   represents the human C1D protein described in the method of the invention
XX
SQ   Sequence 141 AA;

Query Match   90.3%; Score 646; DB 3; Length 141;
Best Local Similarity   90.0%; Pred. No. 5.9e-59;
Matches 126; Conservative   7; Mismatches   7; Indels   0; Gaps   0;

Qy   1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60
          ||||||||||||||||||||||:|::||||||:||||||||||||||||||||||||||||
Db   1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKKAAKLDRGAASRFVKK 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKKAGKLDRGAASRFVKN 120

Qy   121 ALWEPKRKSTPKVANKGKSK 140
          ||||||:|: ||||||||
Db   121 ALWEPKSKNASKVANKGKSK 140

RESULT 4
AD ADE59764
ID ADE59764 standard; protein; 141 AA.
XX
AC ADE59764;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human Protein NP_006324, SEQ ID NO 5660.
XX
KW Human; pain; neuronal tissue; gene therapy; chronic constriction injury; CCI;
KW spinal segmental nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333147P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; NP_006324.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a

---

CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 141 AA;

Query Match   90.3%; Score 646; DB 7; Length 141;
Best Local Similarity   90.0%; Pred. No. 5.9e-59;
Matches 126; Conservative   7; Mismatches   7; Indels   0; Gaps   0;

Qy   1 MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60
          ||||||||||||||||||||||:|::||||||:||||||||||||||||||||||||||||
Db   1 MAGEEINEDYPVEIHEYLSAFENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKKAAKLDRGAASRFVKK 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKKAGKLDRGAASRFVKN 120

Qy   121 ALWEPKRKSTPKVANKGKSK 140
          ||||||:|: ||||||||
Db   121 ALWEPKSKNASKVANKGKSK 140

RESULT 5
AD ADI28083
ID ADI28083 standard; protein; 140 AA.
XX
AC ADI28083;
XX
DT 22-APR-2004 (first entry)
XX
DE Human nuclear receptor cofactor CF6 protein.
XX
KW nuclear receptor cofactor; CF6; cellular function inhibition;
KW metabolic disorder; immunological indication; hormonal dysfunction;
KW neurosystemic disease.
XX
OS Homo sapiens.
XX
PN WO2002224728-A2.
XX
PD 28-MAR-2002.
XX
PF 17-SEP-2001; 2001WO-EP010744.
XX
PR 22-SEP-2000; 2000EP-00120722.
XX
PA (LION-) LION BIOSCIENCE AG.
XX
PI Casari G, Jackson D;
XX
DR WPI; 2002-383179/41.
DR N-PSDB; ADI28090.
XX
PT Polynucleotide and polypeptide of novel nuclear receptor cofactor useful
PT for screening drugs regulating cofactor-associated physiological
PT responses e.g. hormonal dysfunctions.
XX
PS Claim 12; SEQ ID NO 3; 97pp; English.
XX
CC The invention relates to an isolated polynucleotide encoding a nuclear

CC   receptor cofactor, also known as CF6. The polynucleotide or encoded
CC   protein is useful for construction of multiple nuclear receptor cofactor
CC   specific sequence alignments, preferably for the construction of protein
CC   sequence alignments. The protein is useful for screening agents capable
CC   of inhibiting the cellular function of the cofactor CF6. The
CC   polynucleotide is useful for making vectors and for transforming cells,
CC   both of which are ultimately useful for production of the CF6 protein.
CC   They are also useful as scientific research tools for developing nucleic
CC   acid probes for determining expression levels of the cofactor gene, e.g.
CC   to identify diseased or otherwise abnormal states. They are particularly
CC   useful for diagnostic purposes to e.g. identify deleted or mutant CF6
CC   genes; or their measure expression. They are useful for developing
CC   analytical tools such as antisense oligonucleotide for selectively
CC   inhibiting expression of the cofactor gene to determine the cofactor CF7
CC   antagonist activity, and therefore, for screening drugs for agonist and
CC   antagonist. The protein is useful for screening drugs useful in
CC   regulating physiological responses associated with the cofactors such as
CC   metabolic disorders, immunological indications, hormonal dysfunction,
CC   neurosystemic diseases. The proteins are also useful for developing
CC   antibodies for detection of the proteins and are also used to
CC   model the three-dimensional structure of the protein. This sequence
CC   represents the CF6 protein. (Note: this sequence appears to be a fragment
CC   to design primers for a polymerase chain reaction and are also used to
CC   of the CF6 protein as the sequence differs from the CF6 protein sequence
CC   given in Fig 3 (also designated as SEQ ID NO: 2) of the specification).
XX
SQ   Sequence 140 AA;

Query Match   86.0%;   Score 615;   DB 5;   Length 140;
Best Local Similarity   85.7%;   Pred. No. 9.7e-56;
Matches 120;   Conservative 10;   Mismatches 10;   Indels 0;   Gaps 0;

```
Qy   1 MAGEEMNEDYPVEIHESLITALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60
         |||||||||||||:|:|:||||||||||:|:||||||||||||||||||||||||||||
Db   1 MAAEEINEDYPVEIHDYLSAFANSIDAVDEMLKNMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy  61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK 120
         ||||||||||||||||||||||||||||||||||||||||:||||||||||:|||:
Db  61 YTLNSMFVVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVRN 120

Qy 121 ALWEPKRKSTPKVANKGKSK 140
         ||||||:||||:||:||||
Db 121 ALWEPKPKNTSKVAHKGKSK 140
```

RESULT 6
AA015405
ID   AA015405 standard; protein; 141 AA.
XX
AC   AA015405;
XX
DT   27-SEP-2002 (first entry)
XX
DE   Mammalian nuclear receptor cofactor CF7 protein.
XX
KW   Mammalian; nuclear receptor cofactor; CF7; CF8; metabolism regulation;
KW   cell homeostasis; cell proliferation; differentiation;
KW   pathological cellular aberration; cellular defence mechanism.
XX
OS   Mammalia.
XX
PN   WO200242322-A2.
XX
PD   30-MAY-2002.
XX
PF   21-NOV-2001; 2001WO-EP013548.
XX
PR   21-NOV-2000; 2000EP-00125524.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Jackson D, Casari G, Suckow J;
XX

DR   WPI; 2002-566559/60.
DR   N-PSDB; AAL43973.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.
XX
PS   Claim 12; Fig 3; 68pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of two
CC   mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC   CF8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CF7
CC   and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC   involved in cellular functions, such as: regulation of metabolism and
CC   cell homeostasis; cell proliferation and differentiation, pathological
CC   cellular aberrations, or cellular defence mechanisms. The present amino
CC   acid sequence represents the mammalian nuclear receptor cofactor CF7
CC   protein
XX
SQ   Sequence 141 AA;

Query Match   86.0%;   Score 615;   DB 5;   Length 141;
Best Local Similarity   85.7%;   Pred. No. 9.8e-56;
Matches 120;   Conservative 10;   Mismatches 10;   Indels 0;   Gaps 0;

```
Qy   1 MAGEEMNEDYPVEIHESLITALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSA 60
         |||||||||||||:|:|:|||||||||:|:||||||||||||||||||||||||||||
Db   1 MAAEEINEDYPVEIHDYLSAFANSIDAVDEMLKNMMSVSRNELLQKLDPLEQAKVDLVSA 60

Qy  61 YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRFVKK 120
         |||||||||||||||||||||||||||||||||||||||:||||||||||:||||
Db  61 YTLNSMFVVYLATQGVNPKEHPVKQELERIRVYMNRVKEITDKKKAGKLDRGAASRFVRN 120

Qy 121 ALWEPKRKSTPKVANKGKSK 140
         ||||||:||||:||:||||
Db 121 ALWEPKPKNTSKVAHKGKSK 140
```

RESULT 7
AA015406
ID   AA015406 standard; protein; 140 AA.
XX
AC   AA015406;
XX
DT   27-SEP-2002 (first entry)
XX
DE   Mammalian nuclear receptor cofactor CF8 protein.
XX
KW   Mammalian; nuclear receptor cofactor; CF7; CF8; metabolism regulation;
KW   cell homeostasis; cell proliferation; differentiation;
KW   pathological cellular aberration; cellular defence mechanism.
XX
OS   Mammalia.
XX
PN   WO200242322-A2.
XX
PD   30-MAY-2002.
XX
PF   21-NOV-2001; 2001WO-EP013548.
PR   21-NOV-2000; 2000EP-00125524.
XX
PA   (LION-) LION BIOSCIENCE AG.
XX
PI   Jackson D, Casari G, Suckow J;
XX
DR   WPI; 2002-566559/60.
DR   N-PSDB; AAL43975.
XX
PT   Novel nuclear receptor cofactors, CF7 and CF8 for identifying modulators
PT   useful for inhibiting cellular function of cofactor and for treating
PT   metabolic disorders, immunological indications and hormonal dysfunctions.

XX
PS   Claim 12; Fig 3; 68pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of two
CC   mammalian nuclear receptor cofactor proteins (CF7 and CF8). The CF7 and
CC   CF8 protein sequences of the invention are useful for screening agents
CC   that are capable of inhibiting the cellular function of cofactor CF7
CC   and/or CF8. The CF7 and CF8 DNA sequences are useful for modulating genes
CC   involved in cellular functions, such as: regulation of metabolism and
CC   cell homeostasis, cell proliferation and differentiation, pathological
CC   cellular aberrations, or cellular defence mechanisms. The present amino
CC   acid sequence represents the mammalian nuclear receptor cofactor CF8
CC   protein
XX
SQ   Sequence 140 AA;

```
Query Match         81.0%;  Score 579;  DB 5;  Length 140;
Best Local Similarity    82.0%;  Pred. No. 5.4e-52;
Matches 114;  Conservative  9;  Mismatches 16;  Indels  0;  Gaps  0;

Qy      2  AGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMWAVSRNELLQKLDPLEQAKVDLVSAY  61
              |||| |||||||| :  ||  :  ||||:||| ||:||||||||||||||||||||||||
Db      1  AAEEINEDYPVEIHDYLSAPANSIDAVDEMLKNMMSVSRNELLQKLDPLEQAKVDLVSAY  60

Qy     62  TYLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITDKKKAAKLDRGAASRFVKKA  121
              |||||||| |||||||||||| |||||||||||| ||||| ||||| |||||||| ||
Db     61  ILNSMFWAYLATQGVNPKEHSVKQELERIRVYMNRVKEIADKKKAGKLDRGAASRFVRNA  120

Qy    122  LWEPKRKSTPKVANKGKSK  140
              |||| | |||  |||||
Db    121  LWEPKPKNASKVAYKGKSK  139
```

RESULT 8
AAG01231
ID   AAG01231 standard; protein; 102 AA.
XX
AC   AAG01231;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 5312.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;   99US-0122487P.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC01237.
XX
PS   Claim 13; SEQ ID NO 5312; 71pp + Sequence Listing; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number of
CC   5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC   prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC   tissues. EST sequences usually correspond mainly to the 3' untranslated

CC   region (UTR) of the mRNA because they are often obtained from oligo-dT
CC   primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC   sequences derived from the 5' ends of mRNAs and even in those cases where
CC   longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC   included. 5' ESTs are derived from mRNA with intact 5' ends and can
CC   therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC   are also used in diagnostic, forensic, gene therapy and chromosome
CC   mapping procedures. They are used to obtain upstream regulatory sequences
CC   and to design expression and secretion vectors
XX
SQ   Sequence 102 AA;

```
Query Match         67.8%;  Score 485;  DB 3;  Length 102;
Best Local Similarity    92.2%;  Pred. No. 2.1e-42;
Matches 94;  Conservative  6;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1  MAGEEMNEDYPVEIHESLTALESSLGAVDDMLKTMWAVSRNELLQKLDPLEQAKVDLVSA  60
              ||||| |||||||  :  |:|  ||||:||||| ||:|||||||||||||||||||||||
Db      1  MAGEEINEDYPVEIHRYLSAPENSIGAVDEMLKTMMSVSRNELLQKLDPLEQAKVDLVSA  60

Qy     61  YTLNSMFWVYLATQGVNPKEHPVKQELERIRVYMRVKEITD  102
              ||||||||||||||||||| |||||||||||| ||||||
Db     61  YTLNSMFWVYLATQGVNPKEHSVKQELERIRVYMNRVKEITD  102
```

RESULT 9
ABO54730
ID   ABO54730 standard; protein; 54 AA.
XX
AC   ABO54730;
XX
DT   29-JUL-2004 (first entry)
XX
DE   Human genome derived single exon protein #964.
XX
KW   Human; gene expression; single exon probe; microarray;
KW   alternative splicing event; genomic alteration.
XX
OS   Homo sapiens.
XX
PN   US2003194704-A1.
XX
PD   16-OCT-2003.
XX
PF   03-APR-2002; 2002US-00029386.
XX
PR   03-APR-2002; 2002US-00029386.
XX
PA   (PENN/) PENN S G.
PA   (RANK/) RANK D R.
PA   (HANZ/) HANZEL D K.
XX
PI   Penn SG, Rank DR, Hanzel DK;
XX
DR   WPI; 2004-119264/12.
XX
PT   New human genome-derived single exon nucleic acid probes useful for human
PT   gene expression analysis, for identifying or characterizing alternative
PT   splicing event, for assessing genomic alterations or as tools for
PT   surveying tissues.
XX
PS   Claim 45; SEQ ID NO 28364; 80pp; English.
XX
CC   The invention relates to a nucleic acid probe for measuring human gene
CC   expression, comprising any of the 27,400 fully defined nucleotide
CC   sequences in the specification, or their complements or fragments, and
CC   encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC   fully defined in the specification. The probe is a single exon probe that
CC   hybridises under high stringency conditions to a nucleic acid molecule
CC   expressed in human cells or tissues. Also included are a spatially-
CC   addressable set of single exon nucleic acid probes for measuring human
CC   gene expression (comprising a plurality of single exon nucleic acid
CC   probes cited above, where each of the plurality of probes is separately

CC   and addressably isolatable or amplifiable from the plurality), a single
CC   exon microarray for measuring human gene expression, a method of
CC   measuring human gene expression, a vector comprising the single exon
CC   probe cited above, an ORF-encoded peptide comprising at least 8
CC   contiguous amino acids of any of the above- mentioned amino acid
CC   sequences (optionally with conservative amino acid substitutions), an
CC   isolated antibody that binds specifically to a peptide cited above,
CC   methods of selling and/or licensing single exon probes or microarrays to
CC   a customer desiring to measure gene expression, a method of providing
CC   human gene expression data by subscription, and a computer-readable
CC   storage medium which contains a database having a plurality of records
CC   (each record including data on the expression of a single exon probe
CC   cited above. The probe, methods and apparatus are useful in gene
CC   expression analysis. The probes may be used as tools for surveying
CC   specific exon, or in constructing genome-derived single exon microarrays.
CC   In addition, the probes are used in detecting and characterising
CC   tissues to detect the presence of expressed messages that contain their
CC   alterations in the genomic locus that includes their exon, in assessing
CC   alternative splicing events, in identifying and characterising gross
CC   smaller genomic alterations, in priming the synthesis of nucleic acids,
CC   or in expressing the ORF-encoded peptide. The present sequence is a human
CC   single exon probe protein of the invention. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=200301947004
XX
SQ   Sequence 54 AA;

Query Match   32.9%; Score 235; DB 8; Length 54;
Best Local Similarity   88.7%;
Matches   47; Conservative   1; Mismatches   5; Indels   0; Gaps   0;

Qy   88 ERIRVYMNRVKEITDKKKAAKLDRGAASRFVKKALWEPKRKSTPKVANKGKSK 140
                |||||||||||||||||:||||||||||||| |||| | ||||||||||||||
Db   1 ERIRVYMNRVKEITDKKKAGKLDRGAASRFVXNALWEPKPKDVASKVANKGKSK 53

RESULT 10
AAG49051
ID   AAG49051 standard; protein; 217 AA.
XX
AC   AAG49051;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62013.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0129884P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132863P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134205P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135452P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0137282P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
PR   04-JUN-1999;   99US-0137724P.
PR   07-JUN-1999;   99US-0180094P.
PR   08-JUN-1999;   99US-0185540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   18-JUN-1999;   99US-0139460P.
PR   18-JUN-1999;   99US-0139461P.
PR   18-JUN-1999;   99US-0139462P.
PR   18-JUN-1999;   99US-0139463P.
PR   16-JUN-1999;   99US-0139750P.
PR   17-JUN-1999;   99US-0139453P.
PR   18-JUN-1999;   99US-0139763P.
PR   21-JUN-1999;   99US-0139817P.
PR   23-JUN-1999;   99US-0139899P.
PR   23-JUN-1999;   99US-0140353P.
PR   24-JUN-1999;   99US-0140354P.
PR   28-JUN-1999;   99US-0140695P.
PR   29-JUN-1999;   99US-0140823P.
PR   30-JUN-1999;   99US-0140991P.
PR   01-JUL-1999;   99US-0141287P.
PR   01-JUL-1999;   99US-0141842P.
PR   02-JUL-1999;   99US-0142154P.
PR   06-JUL-1999;   99US-0142055P.
PR   08-JUL-1999;   99US-0142390P.
PR   09-JUL-1999;   99US-0142803P.
PR   12-JUL-1999;   99US-0142920P.
PR   13-JUL-1999;   99US-0142977P.
PR   14-JUL-1999;   99US-0143542P.
PR   15-JUL-1999;   99US-0143624P.
PR   16-JUL-1999;   99US-0144005P.
PR   16-JUL-1999;   99US-0144085P.
PR   19-JUL-1999;   99US-0144331P.
PR   19-JUL-1999;   99US-0144325P.
PR   19-JUL-1999;   99US-0144331P.
PR   19-JUL-1999;   99US-0144333P.
PR   19-JUL-1999;   99US-0144334P.
PR   20-JUL-1999;   99US-0144335P.
PR   20-JUL-1999;   99US-0144352P.
PR   20-JUL-1999;   99US-0144632P.
PR   20-JUL-1999;   99US-0148848P.

```
PR   21-JUL-1999;   99US-0144814P.
PR   21-JUL-1999;   99US-0145086P.
PR   21-JUL-1999;   99US-0145088P.
PR   22-JUL-1999;   99US-0145085P.
PR   22-JUL-1999;   99US-0145087P.
PR   22-JUL-1999;   99US-0145089P.
PR   22-JUL-1999;   99US-0145192P.
PR   23-JUL-1999;   99US-0145145P.
PR   23-JUL-1999;   99US-0145218P.
PR   23-JUL-1999;   99US-0145224P.
PR   26-JUL-1999;   99US-0145276P.
PR   27-JUL-1999;   99US-0145913P.
PR   27-JUL-1999;   99US-0145918P.
PR   28-JUL-1999;   99US-0145919P.
PR   28-JUL-1999;   99US-0145951P.
PR   02-AUG-1999;   99US-0146386P.
PR   02-AUG-1999;   99US-0146388P.
PR   02-AUG-1999;   99US-0146389P.
PR   03-AUG-1999;   99US-0147038P.
PR   04-AUG-1999;   99US-0147302P.
PR   05-AUG-1999;   99US-0147192P.
PR   05-AUG-1999;   99US-0147260P.
PR   06-AUG-1999;   99US-0147303P.
PR   09-AUG-1999;   99US-0147416P.
PR   09-AUG-1999;   99US-0147935P.
PR   10-AUG-1999;   99US-0148171P.
PR   11-AUG-1999;   99US-0148319P.
PR   12-AUG-1999;   99US-0148341P.
PR   13-AUG-1999;   99US-0148565P.
PR   13-AUG-1999;   99US-0148684P.
PR   16-AUG-1999;   99US-0149368P.
PR   17-AUG-1999;   99US-0149175P.
PR   18-AUG-1999;   99US-0149426P.
PR   20-AUG-1999;   99US-0149722P.
PR   20-AUG-1999;   99US-0149723P.
PR   20-AUG-1999;   99US-0149929P.
PR   23-AUG-1999;   99US-0149902P.
PR   23-AUG-1999;   99US-0149930P.
PR   25-AUG-1999;   99US-0150566P.
PR   26-AUG-1999;   99US-0150884P.
PR   27-AUG-1999;   99US-0151065P.
PR   27-AUG-1999;   99US-0151066P.
PR   30-AUG-1999;   99US-0151080P.
PR   31-AUG-1999;   99US-0151303P.
PR   01-SEP-1999;   99US-0151438P.
PR   07-SEP-1999;   99US-0151930P.
PR   10-SEP-1999;   99US-0152363P.
PR   13-SEP-1999;   99US-0153070P.
PR   15-SEP-1999;   99US-0153758P.
PR   16-SEP-1999;   99US-0154018P.
PR   20-SEP-1999;   99US-0154039P.
PR   22-SEP-1999;   99US-0154779P.
PR   24-SEP-1999;   99US-0155139P.
PR   28-SEP-1999;   99US-0155486P.
PR   29-SEP-1999;   99US-0155699P.
PR   04-OCT-1999;   99US-0156458P.
PR   05-OCT-1999;   99US-0156596P.
PR   06-OCT-1999;   99US-0157117P.
PR   07-OCT-1999;   99US-0157753P.
PR   08-OCT-1999;   99US-0157865P.
PR   12-OCT-1999;   99US-0158029P.
PR   13-OCT-1999;   99US-0158232P.
PR   13-OCT-1999;   99US-0158369P.
PR   14-OCT-1999;   99US-0159293P.
PR   14-OCT-1999;   99US-0159294P.
PR   14-OCT-1999;   99US-0159295P.
PR   14-OCT-1999;   99US-0159329P.
PR   14-OCT-1999;   99US-0159330P.
PR   14-OCT-1999;   99US-0159331P.
PR   14-OCT-1999;   99US-0159377P.
PR   14-OCT-1999;   99US-0159638P.

PR   18-OCT-1999;   99US-0159568P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   21-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   22-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   25-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   26-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.
```

```
Query Match   21.6%;  Score 154.5;  DB 3;  Length 217;
Best Local Similarity   30.6%;  Pred. No. 1.2e-07;
Matches   41;  Conservative  30;  Mismatches  46;  Indels  17;  Gaps   3;

Qy   22 ESSLGAVDD----------MLKTMMAVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWYL 71
            ||::  ||:         |  :| :|:  :||:|      |:| ::| ::||:||
Db   18 ESAIEAVNQTLAYLKELKPQLEQMLTLAEPEVLAAMQPLQRAKTMHLLAEATTTLYELRL 77

Qy   72 ATQGVNPKEHPVKQELERIRVYMNRVKEITDKKK------AAKLDRGAASRFVKKALWEPK 126
            : : |: |||||  |:||:|:||  | :|||||| |       |:: |::||:::||| ::
Db   78 RCTGVDPDDHRVKSIERINVYREKFQQKCVDQSKGPLRPTTVLNRQAATRFIEHSL--PD 135

Qy  127 RKSTPKVANKGKSK 140
            |::|| : ::||:|
Db  136 LTSTQKQSIRDLSK 149
```

RESULT 11

AAG24127

ID   AAG24127 standard; protein; 217 AA.
XX
AC   AAG24127;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27683.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.

```
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0125488P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0128845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
```

PR 23-APR-1999;   99US-0130991P.
PR 28-APR-1999;   99US-0131449P.
PR 30-APR-1999;   99US-0132048P.
PR 30-APR-1999;   99US-0132407P.
PR 04-MAY-1999;   99US-0132484P.
PR 05-MAY-1999;   99US-0132485P.
PR 06-MAY-1999;   99US-0132486P.
PR 06-MAY-1999;   99US-0132487P.
PR 07-MAY-1999;   99US-0132863P.
PR 11-MAY-1999;   99US-0134256P.
PR 14-MAY-1999;   99US-0134218P.
PR 14-MAY-1999;   99US-0134219P.
PR 14-MAY-1999;   99US-0134221P.
PR 18-MAY-1999;   99US-0134370P.
PR 19-MAY-1999;   99US-0134768P.
PR 20-MAY-1999;   99US-0134941P.
PR 21-MAY-1999;   99US-0135124P.
PR 24-MAY-1999;   99US-0135353P.
PR 25-MAY-1999;   99US-0135629P.
PR 27-MAY-1999;   99US-0136021P.
PR 28-MAY-1999;   99US-0136392P.
PR 01-JUN-1999;   99US-0136782P.
PR 03-JUN-1999;   99US-0137222P.
PR 04-JUN-1999;   99US-0137528P.
PR 07-JUN-1999;   99US-0137502P.
PR 08-JUN-1999;   99US-0137724P.
PR 10-JUN-1999;   99US-0138094P.
PR 10-JUN-1999;   99US-0138540P.
PR 14-JUN-1999;   99US-0138847P.
PR 16-JUN-1999;   99US-0139119P.
PR 16-JUN-1999;   99US-0139452P.
PR 17-JUN-1999;   99US-0139453P.
PR 18-JUN-1999;   99US-0139492P.
PR 18-JUN-1999;   99US-0139454P.
PR 18-JUN-1999;   99US-0139455P.
PR 18-JUN-1999;   99US-0139456P.
PR 18-JUN-1999;   99US-0139457P.
PR 18-JUN-1999;   99US-0139458P.
PR 18-JUN-1999;   99US-0139459P.
PR 18-JUN-1999;   99US-0139460P.
PR 18-JUN-1999;   99US-0139461P.
PR 18-JUN-1999;   99US-0139462P.
PR 18-JUN-1999;   99US-0139463P.
PR 18-JUN-1999;   99US-0139750P.
PR 21-JUN-1999;   99US-0139763P.
PR 22-JUN-1999;   99US-0139817P.
PR 23-JUN-1999;   99US-0139999P.
PR 23-JUN-1999;   99US-0140354P.
PR 24-JUN-1999;   99US-0140353P.
PR 28-JUN-1999;   99US-0140695P.
PR 29-JUN-1999;   99US-0140823P.
PR 30-JUN-1999;   99US-0140991P.
PR 01-JUL-1999;   99US-0141287P.
PR 01-JUL-1999;   99US-0141842P.
PR 02-JUL-1999;   99US-0142154P.
PR 06-JUL-1999;   99US-0142055P.
PR 08-JUL-1999;   99US-0142290P.
PR 12-JUL-1999;   99US-0142003P.
PR 13-JUL-1999;   99US-0142920P.
PR 14-JUL-1999;   99US-0142977P.
PR 15-JUL-1999;   99US-0143542P.
PR 16-JUL-1999;   99US-0143624P.
PR 16-JUL-1999;   99US-0144005P.
PR 19-JUL-1999;   99US-0144085P.
PR 19-JUL-1999;   99US-0144086P.
PR 19-JUL-1999;   99US-0144325P.
PR 19-JUL-1999;   99US-0144331P.
PR 19-JUL-1999;   99US-0144332P.
PR 19-JUL-1999;   99US-0144333P.
PR 20-JUL-1999;   99US-0144334P.
PR 20-JUL-1999;   99US-0144335P.
PR 20-JUL-1999;   99US-0144352P.
PR 20-JUL-1999;   99US-0144632P.

PR 20-JUL-1999;   99US-0144884P.
PR 21-JUL-1999;   99US-0144814P.
PR 21-JUL-1999;   99US-0145086P.
PR 21-JUL-1999;   99US-0145085P.
PR 22-JUL-1999;   99US-0145088P.
PR 22-JUL-1999;   99US-0145087P.
PR 22-JUL-1999;   99US-0145089P.
PR 22-JUL-1999;   99US-0145192P.
PR 23-JUL-1999;   99US-0145145P.
PR 23-JUL-1999;   99US-0145218P.
PR 23-JUL-1999;   99US-0145224P.
PR 26-JUL-1999;   99US-0145276P.
PR 27-JUL-1999;   99US-0145913P.
PR 27-JUL-1999;   99US-0145918P.
PR 28-JUL-1999;   99US-0145919P.
PR 02-AUG-1999;   99US-0145951P.
PR 02-AUG-1999;   99US-0146386P.
PR 02-AUG-1999;   99US-0146388P.
PR 03-AUG-1999;   99US-0146389P.
PR 04-AUG-1999;   99US-0147038P.
PR 04-AUG-1999;   99US-0147204P.
PR 05-AUG-1999;   99US-0147303P.
PR 05-AUG-1999;   99US-0147192P.
PR 06-AUG-1999;   99US-0147260P.
PR 06-AUG-1999;   99US-0174162P.
PR 09-AUG-1999;   99US-0147935P.
PR 09-AUG-1999;   99US-0174993P.
PR 10-AUG-1999;   99US-0181719P.
PR 11-AUG-1999;   99US-0148319P.
PR 12-AUG-1999;   99US-0148341P.
PR 13-AUG-1999;   99US-0148684P.
PR 13-AUG-1999;   99US-0148565P.
PR 16-AUG-1999;   99US-0149368P.
PR 17-AUG-1999;   99US-0149175P.
PR 18-AUG-1999;   99US-0149266P.
PR 20-AUG-1999;   99US-0149722P.
PR 20-AUG-1999;   99US-0149723P.
PR 23-AUG-1999;   99US-0149929P.
PR 23-AUG-1999;   99US-0149902P.
PR 25-AUG-1999;   99US-0149930P.
PR 26-AUG-1999;   99US-0150566P.
PR 27-AUG-1999;   99US-0150884P.
PR 27-AUG-1999;   99US-0151065P.
PR 30-AUG-1999;   99US-0151066P.
PR 31-AUG-1999;   99US-0151080P.
PR 01-SEP-1999;   99US-0151303P.
PR 07-SEP-1999;   99US-0151438P.
PR 10-SEP-1999;   99US-0151930P.
PR 13-SEP-1999;   99US-0152363P.
PR 15-SEP-1999;   99US-0153070P.
PR 16-SEP-1999;   99US-0153758P.
PR 20-SEP-1999;   99US-0154018P.
PR 22-SEP-1999;   99US-0154039P.
PR 23-SEP-1999;   99US-0154779P.
PR 24-SEP-1999;   99US-0155139P.
PR 28-SEP-1999;   99US-0155486P.
PR 29-SEP-1999;   99US-0155659P.
PR 04-OCT-1999;   99US-0156458P.
PR 05-OCT-1999;   99US-0156596P.
PR 06-OCT-1999;   99US-0157117P.
PR 07-OCT-1999;   99US-0157553P.
PR 08-OCT-1999;   99US-0157865P.
PR 12-OCT-1999;   99US-0158029P.
PR 13-OCT-1999;   99US-0158322P.
PR 13-OCT-1999;   99US-0158369P.
PR 13-OCT-1999;   99US-0192293P.
PR 14-OCT-1999;   99US-0192294P.
PR 14-OCT-1999;   99US-0192295P.
PR 14-OCT-1999;   99US-0193330P.
PR 14-OCT-1999;   99US-0193331P.
PR 14-OCT-1999;   99US-0196378P.

RESULT 12

AAG49050

ID   AAG49050 standard; protein; 253 AA.

XX
AC   AAG49050;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62012.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX

Query Match   41; Conservative   30; Mismatches 46; Indels 17; Gaps   3;
Best Local Similarity   30.6%; Pred. No. 1.2e-07;
Matches   41; Conservative   30; Mismatches 46; Indels 17; Gaps   3;

Qy   22 ESSLGAVDD----------------MLKTMMAVSRNELLQKLDPLEQAKVDLVSAYTLNSMFVVYL 71
                            : :|: | | |:||| :|||:|: :| | |||:
Db   18 ESAIEAVNQTLAYLKELKPQLBQMLTLAEPEVLAAMQPLQRAKTMILLAEATTLIYELRL 77

Qy   72 ATQGVNPKEHPVKQELERIRVYMNRVKEITDKKK---------AAKLDRGAAASRFVKKALWEPK 126
               |:|| | ||||:| ||:|: | | | : | :| || | |:
Db   78 RCTGVDPDDHRVKSEIERINVYREKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSL--PD 135

Qy   127 RKSTPKVANKGKSK 140
               || |: || ||
Db   136 LTSTQKQSIRDLSK 149

PR   14-OCT-1999;   99US-0159638P.
PR   18-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160880P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.

PR   25-FEB-1999;   99US-0118825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.

PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132663P.
PR   11-MAY-1999;   99US-0133256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135553P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137228P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
PR   10-JUN-1999;   99US-0138540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   16-JUN-1999;   99US-0139452P.
PR   16-JUN-1999;   99US-0139453P.
PR   17-JUN-1999;   99US-0139492P.
PR   18-JUN-1999;   99US-0139454P.
PR   18-JUN-1999;   99US-0139455P.
PR   18-JUN-1999;   99US-0139456P.
PR   18-JUN-1999;   99US-0139457P.
PR   18-JUN-1999;   99US-0139458P.
PR   18-JUN-1999;   99US-0139459P.
PR   18-JUN-1999;   99US-0139460P.
PR   18-JUN-1999;   99US-0139461P.
PR   18-JUN-1999;   99US-0139462P.
PR   18-JUN-1999;   99US-0139463P.
PR   18-JUN-1999;   99US-0139750P.
PR   18-JUN-1999;   99US-0139763P.
PR   21-JUN-1999;   99US-0139817P.
PR   22-JUN-1999;   99US-0139899P.
PR   23-JUN-1999;   99US-0140353P.
PR   23-JUN-1999;   99US-0140354P.
PR   24-JUN-1999;   99US-0140695P.
PR   28-JUN-1999;   99US-0140823P.
PR   29-JUN-1999;   99US-0140991P.
PR   30-JUN-1999;   99US-0141287P.
PR   01-JUL-1999;   99US-0141842P.
PR   01-JUL-1999;   99US-0142154P.
PR   02-JUL-1999;   99US-0142205P.
PR   06-JUL-1999;   99US-0142390P.
PR   08-JUL-1999;   99US-0142803P.
PR   09-JUL-1999;   99US-0142920P.
PR   12-JUL-1999;   99US-0142977P.
PR   13-JUL-1999;   99US-0143542P.
PR   14-JUL-1999;   99US-0143624P.
PR   15-JUL-1999;   99US-0144005P.
PR   16-JUL-1999;   99US-0144086P.
PR   16-JUL-1999;   99US-0144325P.
PR   19-JUL-1999;   99US-0144331P.
PR   19-JUL-1999;   99US-0144332P.
PR   19-JUL-1999;   99US-0144333P.
PR   19-JUL-1999;   99US-0144335P.
PR   20-JUL-1999;   99US-0144352P.

PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145524P.
PR 26-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145951P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 12-AUG-1999; 99US-0148319P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 17-AUG-1999; 99US-0149368P.
PR 18-AUG-1999; 99US-0149175P.
PR 20-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151438P.
PR 07-SEP-1999; 99US-0151930P.
PR 10-SEP-1999; 99US-0152363P.
PR 13-SEP-1999; 99US-0153070P.
PR 15-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 20-SEP-1999; 99US-0154039P.
PR 22-SEP-1999; 99US-0154779P.
PR 23-SEP-1999; 99US-0155139P.
PR 24-SEP-1999; 99US-0155486P.
PR 28-SEP-1999; 99US-0155659P.
PR 29-SEP-1999; 99US-0156458P.
PR 04-OCT-1999; 99US-0156596P.
PR 05-OCT-1999; 99US-0157117P.
PR 06-OCT-1999; 99US-0157753P.
PR 07-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 09-MAR-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.

PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161360P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match   21.6%; Score 154.5; DB 3; Length 253;
Best Local Similarity   30.6%; Pred. No. 1.5e-07;
Matches   41; Conservative   30; Mismatches   46; Indels   17; Gaps   3;

Qy   22 ESSLGAVDD----------MLKTMMAVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYL 71
          |::| ||: |: ||:. |:||::|: : |||::||:.| |:|: ::.
Db   54 ESAIEAVNQTLAYLKELKPQLEQMLTLAEPEVLAAAMQPLQRAKTMHLLAEATTTLYELRL 113

Qy   72 ATQGVNPKEHPVKQELERIRVYMNRVKEITDKKK------AAKLDRGAASRFVKKALWEPK 126
          . |:.|| : |:||:|||||.||| :| |: ||:|| |||. :|| |:
Db   114 RCTGVDPDDHRVKSEIERINVREKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSL--PD 171

Qy   127 RKSTPKVANKGKSK 140
          |:|| |: |
Db   172 LTSTQKQSIRDLSK 185

RESULT 13
AAG24126
ID   AAG24126 standard; protein; 256 AA.
XX
AC   AAG24126;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27682.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.

PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0147303P.
PR 02-AUG-1999; 99US-0143866P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0171192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0174416P.
PR 09-AUG-1999; 99US-0174932P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148684P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153587P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0155458P.
PR 29-SEP-1999; 99US-0156696P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.

RESULT 14
AAG24128

ID   AAG24128 standard; protein; 177 AA.
XX
AC   AAG24128;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 27684.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999; 2000EP-00301439.
PR   05-MAR-1999; 99US-0121825P.
PR   09-MAR-1999; 99US-0123180P.
PR   23-MAR-1999; 99US-0123548P.
PR   25-MAR-1999; 99US-0125788P.
PR   29-MAR-1999; 99US-0126264P.
PR   01-APR-1999; 99US-0126785P.
PR   06-APR-1999; 99US-0127462P.
PR   08-APR-1999; 99US-0128234P.
PR   16-APR-1999; 99US-0128714P.
PR   19-APR-1999; 99US-0129845P.

Query Match   21.6%; Score 154.5; DB 3; Length 256;
Best Local Similarity   30.6%; Pred. No. 1.5e-07;
Matches   41; Conservative   30; Mismatches   46; Indels   17; Gaps   3;

Qy   22 ESSLGAVDD----------MLKTMAVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYL 71
          |||:|| |||. |:|::| |:|:|| :||::|:|:| ::|
Db   57 ESAIEAVNQTLAYLKELKPQLEQMLTLAEPEVLAAMQPLQRAKTMHLLAEATTTLYELRL 116

Qy   72 ATQGVNPKEHPVKQELERIRVYMRVKEITDKKK---------AAKLDRGAASRFVKKALWEPK 126
          ||:| :|| ::|||:||:|:: | ::|:: |.||: ||:|: |
Db   117 RCTGVDPDDHRVKSEIERINVYREKPQKCVDQSKGPLRPTTVLNRQAATRFIEHSL--PD 174

Qy   127 RKSTPKVANKGKSK 140
          | | |:| ||
Db   175 LITSTQKQSIRDLSK 188

RESULT 15
AAG49052

ID   AAG49052 standard; protein; 177 AA.
AC   AAG49052;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62014.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX

Query Match   20.5%; Score 146.5; DB 3; Length 177;
Best Local Similarity   32.4%; Pred. No. 6.5e-07;
Matches   36; Conservative 26; Mismatches 42; Indels 7; Gaps 2;

Qy   35 MMAVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLIATQGVNPKEHPVKQELERIRVYM 94
              |.:   : :|:.|:|:|| :: .|. | ::|: ||:|| |:|:|:||:||
Db   1 MLTLAEPEVLAAMQPLQRAKTMHLLAEATTTLYELRLRCTGVDPDDHRVKSEIERINVYR 60

Qy   95 NRVKEITDKKK------AAKLDRGAASRFVKKALWBPKRKSTPKVANKGKSK 140
              :: .| :.|. ||| :|.||:||:   : | | ||.: :.|||.|:||:||| |
Db   61 EKFQKCVDQSKGPLRPPTVLNRQAATRFIEHSL--PDLTSTCQKQSIRDLSK 109

PR 23-APR-1999;   99US-0130891P,
PR 28-APR-1999;   99US-0131449P,
PR 30-APR-1999;   99US-0132048P,
PR 30-APR-1999;   99US-0132407P,
PR 04-MAY-1999;   99US-0132484P,
PR 05-MAY-1999;   99US-0132485P,
PR 06-MAY-1999;   99US-0132486P,
PR 06-MAY-1999;   99US-0132487P,
PR 07-MAY-1999;   99US-0132863P,
PR 11-MAY-1999;   99US-0134256P,
PR 14-MAY-1999;   99US-0134258P,
PR 14-MAY-1999;   99US-0134219P,
PR 14-MAY-1999;   99US-0134221P,
PR 18-MAY-1999;   99US-0134370P,
PR 19-MAY-1999;   99US-0134768P,
PR 20-MAY-1999;   99US-0134941P,
PR 21-MAY-1999;   99US-0135124P,
PR 24-MAY-1999;   99US-0135353P,
PR 25-MAY-1999;   99US-0135629P,
PR 27-MAY-1999;   99US-0136021P,
PR 28-MAY-1999;   99US-0136392P,
PR 01-JUN-1999;   99US-0136782P,
PR 03-JUN-1999;   99US-0137222P,
PR 04-JUN-1999;   99US-0137528P,
PR 07-JUN-1999;   99US-0137502P,
PR 08-JUN-1999;   99US-0137724P,
PR 10-JUN-1999;   99US-0138094P,
PR 14-JUN-1999;   99US-0138540P,
PR 16-JUN-1999;   99US-0138847P,
PR 17-JUN-1999;   99US-0139119P,
PR 18-JUN-1999;   99US-0139452P,
PR 18-JUN-1999;   99US-0139453P,
PR 18-JUN-1999;   99US-0139454P,
PR 18-JUN-1999;   99US-0139455P,
PR 18-JUN-1999;   99US-0139456P,
PR 18-JUN-1999;   99US-0139457P,
PR 18-JUN-1999;   99US-0139458P,
PR 18-JUN-1999;   99US-0139459P,
PR 18-JUN-1999;   99US-0139460P,
PR 18-JUN-1999;   99US-0139461P,
PR 18-JUN-1999;   99US-0139462P,
PR 18-JUN-1999;   99US-0139463P,
PR 21-JUN-1999;   99US-0139763P,
PR 22-JUN-1999;   99US-0139817P,
PR 22-JUN-1999;   99US-0138899P,
PR 23-JUN-1999;   99US-0140353P,
PR 23-JUN-1999;   99US-0140354P,
PR 24-JUN-1999;   99US-0140695P,
PR 28-JUN-1999;   99US-0140823P,
PR 29-JUN-1999;   99US-0140991P,
PR 30-JUN-1999;   99US-0141287P,
PR 01-JUL-1999;   99US-0141842P,
PR 01-JUL-1999;   99US-0141154P,
PR 02-JUL-1999;   99US-0142055P,
PR 06-JUL-1999;   99US-0142390P,
PR 08-JUL-1999;   99US-0142803P,
PR 09-JUL-1999;   99US-0142920P,
PR 12-JUL-1999;   99US-0142977P,
PR 13-JUL-1999;   99US-0143542P,
PR 14-JUL-1999;   99US-0143624P,
PR 15-JUL-1999;   99US-0144005P,
PR 16-JUL-1999;   99US-0144085P,
PR 16-JUL-1999;   99US-0144086P,
PR 19-JUL-1999;   99US-0144325P,
PR 19-JUL-1999;   99US-0144331P,
PR 19-JUL-1999;   99US-0144332P,
PR 19-JUL-1999;   99US-0144333P,
PR 19-JUL-1999;   99US-0144334P,
PR 19-JUL-1999;   99US-0144335P,
PR 20-JUL-1999;   99US-0144352P,
PR 20-JUL-1999;   99US-0144632P,

PR 20-JUL-1999;   99US-0144884P,
PR 21-JUL-1999;   99US-0144814P,
PR 21-JUL-1999;   99US-0145086P,
PR 22-JUL-1999;   99US-0145085P,
PR 22-JUL-1999;   99US-0145087P,
PR 22-JUL-1999;   99US-0145089P,
PR 22-JUL-1999;   99US-0145192P,
PR 23-JUL-1999;   99US-0145145P,
PR 23-JUL-1999;   99US-0145218P,
PR 23-JUL-1999;   99US-0145224P,
PR 26-JUL-1999;   99US-0145276P,
PR 27-JUL-1999;   99US-0145913P,
PR 27-JUL-1999;   99US-0145918P,
PR 28-JUL-1999;   99US-0145919P,
PR 27-JUL-1999;   99US-0145951P,
PR 02-AUG-1999;   99US-0146386P,
PR 02-AUG-1999;   99US-0146388P,
PR 02-AUG-1999;   99US-0146389P,
PR 03-AUG-1999;   99US-0147038P,
PR 04-AUG-1999;   99US-0147032P,
PR 04-AUG-1999;   99US-0147302P,
PR 05-AUG-1999;   99US-0147192P,
PR 05-AUG-1999;   99US-0147260P,
PR 06-AUG-1999;   99US-0147303P,
PR 06-AUG-1999;   99US-0147416P,
PR 09-AUG-1999;   99US-0147493P,
PR 09-AUG-1999;   99US-0147935P,
PR 10-AUG-1999;   99US-0148171P,
PR 11-AUG-1999;   99US-0148319P,
PR 12-AUG-1999;   99US-0148341P,
PR 13-AUG-1999;   99US-0148684P,
PR 13-AUG-1999;   99US-0148565P,
PR 16-AUG-1999;   99US-0149368P,
PR 17-AUG-1999;   99US-0149175P,
PR 18-AUG-1999;   99US-0149426P,
PR 20-AUG-1999;   99US-0149722P,
PR 20-AUG-1999;   99US-0149723P,
PR 23-AUG-1999;   99US-0149929P,
PR 23-AUG-1999;   99US-0149902P,
PR 25-AUG-1999;   99US-0149930P,
PR 26-AUG-1999;   99US-0150566P,
PR 27-AUG-1999;   99US-0150884P,
PR 27-AUG-1999;   99US-0151065P,
PR 30-AUG-1999;   99US-0151066P,
PR 31-AUG-1999;   99US-0151080P,
PR 01-SEP-1999;   99US-0151303P,
PR 07-SEP-1999;   99US-0151438P,
PR 10-SEP-1999;   99US-0151930P,
PR 13-SEP-1999;   99US-0152363P,
PR 15-SEP-1999;   99US-0153070P,
PR 16-SEP-1999;   99US-0153758P,
PR 20-SEP-1999;   99US-0154018P,
PR 22-SEP-1999;   99US-0154039P,
PR 23-SEP-1999;   99US-0154779P,
PR 24-SEP-1999;   99US-0155119P,
PR 28-SEP-1999;   99US-0155486P,
PR 29-SEP-1999;   99US-0155659P,
PR 04-OCT-1999;   99US-0156458P,
PR 05-OCT-1999;   99US-0156596P,
PR 06-OCT-1999;   99US-0157117P,
PR 07-OCT-1999;   99US-0157753P,
PR 08-OCT-1999;   99US-0157865P,
PR 12-OCT-1999;   99US-0158029P,
PR 13-OCT-1999;   99US-0158232P,
PR 13-OCT-1999;   99US-0158369P,
PR 14-OCT-1999;   99US-0159294P,
PR 14-OCT-1999;   99US-0152994P,
PR 14-OCT-1999;   99US-0159295P,
PR 14-OCT-1999;   99US-0159329P,
PR 14-OCT-1999;   99US-0159330P,
PR 14-OCT-1999;   99US-0159331P,
PR 14-OCT-1999;   99US-0159637P,

```
PR   14-OCT-1999;   99US-0159638P.
PR   18-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.
```

```
Query Match         20.5%;   Score 146.5;   DB 3;   Length 177;
Best Local Similarity   32.4%;   Pred. No. 6.5e-07;
Matches   36;   Conservative   26;   Mismatches   42;   Indels   7;   Gaps   2;

Qy     35   MMAVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELRIRVYM  94
                 |: :|   |:|  |||:||  :::: :   |||::|  :||||||||||
Db      1   MLTLABPEVLAAMQPLQRAKTMHLLAEATTLYELRLRCTGVDPDDHRVKSEIERINVYR  60

Qy     95   NRVKEITDKKK-------AAKLDRGAASRFVKKALWEPKRKSTPKVANKGKSK  140
                : ::        |:|| |:|:  |: :   |||: : |:
Db     61   EKFQKCVDQSKGPLRPTTVLNRQAATRFIEHSL--PDLTSTOKQSIRDLSK  109
```

Search completed: January 4, 2005, 13:42:28
Job time : 70 secs

This Page Blank (uspto)

OM protein - protein search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: January 4, 2005, 11:20:50

Title: US-09-701-618A-4
Perfect Score: 715
Sequence: 1 MAGEEMNEDYPVEIHESLTA..........LWEPKRKSTPKVANKGKSKH 141
Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 9621676 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR 79:*
               1: pir1:*
               2: pir2:*
               3: pir3:*
               4: pir4:*

               Search time 16.5 Seconds
               (without alignments)
               822.216 Million cell updates/sec

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	ID	Description
1	142	19.9	133	2	T41115	sun-cor hormone re
2	100	14.0	184	2	S46808	hypothetical prote
3	86.5	12.1	508	2	S73430	glycerol kinase gl
4	84.5	11.8	1249	2	H71404	hypothetical prote
5	83.5	11.7	1055	2	H64577	type I restriction
6	82.5	11.5	306	2	T27985	hypothetical prote
7	82.5	11.5	2285	2	T12796	probable transglyc
8	82	11.5	530	2	G71157	hypothetical prote
9	81	11.3	207	2	S70534	hypothetical prote
10	81	11.3	207	2	S70533	bbk2.10 protein pr
11	81	11.3	478	2	C59095	bbk2.10 protein pr
12	80.5	11.3	1133	2	T22976	hypothetical prote
13	80	11.2	317	2	T05528	hypothetical prote
14	80	11.2	992	2	S54396	protein L precurso
15	79	11.0	224	2	E64542	conserved hypothet
16	79	11.0	872	2	C72042	CT619 hypothetical
17	79	11.0	872	2	D86581	hypothetical prote
18	78.5	11.0	250	2	T22093	hypothetical prote
19	78.5	11.0	415	1	A48359	glutamyl-tRNA redu
20	78	10.9	550	2	B71441	probable myosin II
21	77	10.8	304	2	F72346	M-related protein
22	77	10.8	444	2	S64912	probable membrane
23	77	10.8	478	2	S25821	transposase - Baci
24	76.5	10.7	649	2	C75112	hypothetical prote
25	76.5	10.7	1134	2	D75014	hypothetical prote
26	76.5	10.7	1156	2	B70356	chromosome assembl
27	76.5	10.7	1708	2	AE1866	WD-40 repeat prote
28	76.5	10.7	1992	2	A47297	myosin heavy chain
29	76	10.6	471	2	S76021	probable trigger f
30	76	10.6	2415	1	A33733	spectrin alpha cha
31	75.5	10.6	117	2	T14973	hypothetical prote
32	75.5	10.6	604	2	T19682	hypothetical prote
33	75.5	10.6	629	2	F84428	probable myosin he
34	75.5	10.6	635	2	D86265	hypothetical prote
35	75.5	10.6	648	2	G72279	ATP-dependent DNA
36	75.5	10.6	2094	2	S33124	tpr protein - huma
37	75	10.5	295	2	AB2464	tRNA delta-2-isope
38	75	10.5	755	2	T20320	hypothetical prote
39	75	10.5	791	2	T20815	DNA repair protein
40	74.5	10.4	326	2	C64480	DNA repair protein
41	74.5	10.4	375	2	T32251	hypothetical prote
42	74.5	10.4	441	2	B71816	hypothetical prote
43	74.5	10.4	481	2	T28900	hypothetical prote
44	74.5	10.4	670	2	D86457	hypothetical prote
45	74.5	10.4	970	2	D59435	Gem-interacting pr

ALIGNMENTS

RESULT 1
T41115
sun-cor hormone receptor repression complex homolog - fission yeast (Schizosaccharomyces
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T41115
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21969
A;Accession: T41115
A;Molecule type: DNA
A;Status: Preliminary; translated from GB/EMBL/DDBJ
A;Experimental source: strain 972h-; cosmid c1739
A;Cross-references: UNIPROT:O74469; EMBL:AL031540; PIDN:CAA20781.1; GSPDB:GN00068; SPDB:S
C;Genetics:
A;Genetics: SPDB:SPCC1739.07
A;Map position: 3

Query Match 19.9%; Score 142; DB 2; Length 133;
Best Local Similarity 28.6%; Conservative 24; Mismatches 57; Indels 14; Gaps 3;
Matches 38;

Qy   13 EIHESLTALESSIGAVDDMLKTMM-AVSRNELLQKLDPLEQAKVDLVSAYTLNSMFWVYL 71
             |:|||:|||:||||.:||.||| |||:.:|.::|:.||:|.||..|.||.||.:|||:|
Db   4 EYSELFERLNRQLDNVEDVLKPLKDAESIFELAEGKSELEQAKLYITMSYAINSTLYSFY 63

Qy   72 ATQGVNPKEHPVKQELERIRVYMNRV----KEITDKKKAAKLDRGAASRFVKKALWEPKR 127
             :|| |||.|:.||.||.|:|.||:. |:||.||.|.||.|||.|:|.||.||||
Db   64 KLNGIDASERPVMQELQRVKNYISKIQQAEKNVNPKTEAVNTSNAIS------SSS 114

Qy   128 KSTPKVANKGKSK 140
            :||||||||||
Db   115 SNRPKVAKDAATR 127

RESULT 2
S46808
hypothetical protein YHR081w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jul-2004
C;Accession: S46808
R;Favello, T.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of S. cerevisiae cosmid 9205.
A;Reference number: S46795
A;Accession: S46808
A;Molecule type: DNA
A;Residues: 1-184 <FAV>
A;Cross-references: UNIPROT:P38801; EMBL:U10556; NID:g500825; PID:g500829; GSPDB:GN00008,
C;Genetics:
A;Gene: MIPS:YHR081w

A;Cross-references: SGD:S000l123
A;Map position: 8R

Query Match   14.0%;   Score 100; DB 2; Length 184;
Best Local Similarity   24.8%;   Pred. No. 0.2;
Matches 31; Conservative 34; Mismatches 48; Indels 12; Gaps 6;

Qy   27 AVDDM---LKTMMAVSRNELLQKLDPLEQAKVDLVS--AYTLNSMFWVYLATQGVNPKEH   81
                |:: :|| || |: ||| ||| || : || :: |::|| | | : | : |
Db   17 ALDELKPEIEKLTSKSLDEQLLLLSD-ERAKLELINRYAYVLVLSSLMFANMKVLGVKDMS- 74

Qy   82 PVKQELERIRVYMNRVKE---ITDKKAAKLDRGAASRFVKKALWEPRKRKSTPKVANKG   137
                ::::: :: ||| || :: || : |||| |: : ||: | | |: : ::
Db   75 PILGELKRVKSYMDKAKQYDNRITKSNEKSQABQEKAKNIISNVLDGNKNQQFPSISRSN   134

Qy   138 -KSKH 141
            |:|
Db   135 FQGKH 139

RESULT 3
S73430
glycerol kinase glpK - Mycoplasma pneumoniae (strain ATCC 29342)
N;Alternate names: hypothetical protein D09_orf508
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
A;Molecule type: DNA
C;Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S73430
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A;Reference number: 24, 4420-4449, 1996
Nucleic Acids Res.
A;Cross-references: UNIPROT:P75064; EMBL:AB000012; GB:U00089; NID:g1673755; PIDN:AAB95757
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Residues: 1-508 <HIM>
A;Gene: glpK
A;Genetics:
C;Genetics:
A;Genetic code: SGC3
C;Superfamily: xylulokinase

Query Match   12.1%;   Score 86.5; DB 2; Length 508;
Best Local Similarity   26.4%;   Pred. No. 9;
Matches 33; Conservative 20; Mismatches 53; Indels 19; Gaps 5;

Qy   10 YPVEIHESLTALESSL-----GA-----VDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSAYT   62
                : ::: :|| : : || |::: || || ||||| || || :: :| :
Db   298 WOLENQKPVYALEGGSVFVAGAALKWLRDSLKVMYSAAESDFYAKLALQKEEQEVFVPAFT 357

Qy   63 -LNSMFW------VYLATQGVNPKEHPVKQELERIRVYMN-----RVKEITDKKKVAAKLD 110
             ||: | | || | |: | :: || |:| | : | ::: || || |||
Db   358 GLGAPYWDASARGAIFGIEANTKREHLVKATLEAIAFQANDLIKAMASDLNSSIKKIKAD 417

Qy   111 RGAAS 115
           || |
Db   418 GGACN 422

RESULT 4
H71404
hypothetical protein dl3195c - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 09-Jul-2004
C;Accession: H71404
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans

C.; Chalwatzig, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A;Reference number: H71404
A;Accession: H71404
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1249 <BEV>
A;Cross-references: UNIPROT:O23289; GB:Z97335; NID:g2244747; PID:g2244787
C;Genetics:
A;Map position: 4COP9-4G3845
C;Superfamily: Arabidopsis thaliana hypothetical protein dl3195c

Query Match   11.8%;   Score 84.5; DB 2; Length 1249;
Best Local Similarity   21.8%;   Pred. No. 37;
Matches 34; Conservative 31; Mismatches 60; Indels 31; Gaps 5;

Qy   5 EMNED-----------------YPVEIHESLTALESSLGAV-DDMLKTMMAVSRN--------- 41
                                  ||:|::: : || || || | | || ||::
Db   277 EMSKDIRGNEGGSSNSTAKYPSKLHEKLAFLEGKVKKIASDIKKTKDMLDLNNPDSSKVII 336

Qy   42 -ELLQKLDPLEQAKVDLVSAYTLNSMFWVYLATQGVNPKEHPVKQELE-RIRVYMNRVKE   99
            :|| ||: || :| |::   :|: | | | || || :| || || :|:|:| | ||
Db   337 SDIHQKITGIEKSMSHVIDGPEKNKTTKAKSSVKGLN-----KEELEDRLLPHQRLLRS 390

Qy   100 ITDKKAAKLDRGAASRFVKKALWEPRKRKSTPKVAN 135
           :| || | |: | :: : : | |: | | |
Db   391 RTQSKITSSHVSKGHDSVESNKAVNAEEKPSAPVEEN 426

RESULT 5
H64577
type I restriction enzyme R protein - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: H64577
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: H64577
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1055 <TOM>
A;Cross-references: UNIPROT:O25211; GB:AE000561; GB:AE000511; NID:g2313564; PIDN:AAD07526
A;Cross-references: UNIPROT:O25211; GB:AE000561; GB:AE000511; NID:g2313564; PIDN:AAD07526

Query Match   11.7%;   Score 83.5; DB 2; Length 1055;
Best Local Similarity   23.4%;   Pred. No. 37;
Matches 30; Conservative 32; Mismatches 37; Indels 29; Gaps 6;

Qy   9 DYPVEIHESLTALESSLGAVDDMLKTMMAVS-----RNELL-----QKLDPLEQAKVDLVS   59
           :|| | ::: : ||: | :: ||: ||:| |: || |||| || :| |:
Db   764 DYPIDDIEGMTSAIVSMSAMNELVKVSRAINTLKERVNLIRTSNDKKILSLKB-KIDIEK 822

Qy   60 AYTLNSMFWVYL-----ATQGVNPKEHP----VKQEL------------ERIRVYMNRVKEI 100
           | | ||: |: | || |::: | ||:|:   :|:| | | ::
Db   823 IHKISSMLHQKAKHLHALKNINEPKPNDLMILEDLIALLDFKIEFKERKELRFKEQEEI 882

Qy   101 TDKKKAAK 108
           |||| ||
Db   883 TTKQKQAK 890

RESULT 6
T27985
hypothetical protein ZK721.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27985
R;Miller, N.

submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid ZK721.
A;Reference number: Z20450
A;Accession: T27985
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-306 <MIL>
A;Cross-references: UNIPROT:Q9GYP1; EMBL:U40951; PIDN:AAA81735.1; CESP:ZK721.2
C;Genetics:
A;Gene: CESP:ZK721.2
A;Introns: 37/3; 64/3; 114/3; 205/2; 252/3; 296/1
C;Superfamily: troponin I

```
Query Match           11.5%; Score 82.5; DB 2; Length 306;
Best Local Similarity 26.0%; 
Matches  40; Conservative  27; Mismatches  62; Indels  25; Gaps  7;

Qy    2 AGEMNEDYPVEIHESLTALES------SLGAVDDMLKTMAVSRNELLQKLDPLEQAKV 55
Db  121 AAEDLKROQLKEQERQKALADRTISLPNVDSIDD--KGQLEKIVNDLWARLTQLEBEKY 178

Qy   56 DL--VSAYT---LNSMFWVYLATQGSVNP-KEHPVKQELERIRVYMNRVKEITDKKKAAKL 109
Db  179 DINVVSQTEAEINSL------TIEVNDLRGKFVKPSLKKVSKYDNKFKKSGBSKAGTKE 232

Qy  110 DRGAASRFVKKALWE-----------PKRSTPKVANKGK 138
Db  233 DFRANLKIVKKDVMEAIVNVKKDDKPDWSKKKNK 266
```

RESULT 7
T12796
probable transglycosylase - Bacillus subtilis phage SPbc2
C;Species: Bacillus subtilis phage SPbc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: T12796; A6991I
R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
A;Description: The complete nucleotide sequence of the Bacillus subtilis Spbeta2 propha
A;Reference number: 217583
A;Accession: T12796
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2285 <LAZ>
A;Cross-references: UNIPROT:O64046; EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AAC1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
A;Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A;Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69911
A;Accession: A69911
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2285 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1; PID:e1183582;
A;Experimental source: strain 168

```
Query Match           11.5%; Score 82.5; DB 2; Length 2285;
Best Local Similarity 22.3%; Pred. No. 1.1e-02;
Matches  33; Conservative  31; Mismatches  53; Indels  31; Gaps  5;
```

RESULT 8
G71157
hypothetical protein PH0462 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C;Accession: G71157
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: G71157
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-530 <KAW>
A;Cross-references: UNIPROT:O58213; GB:AP000002; NID:g3236129; PIDN:BAA29548.1; PID:g325
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0462

```
Query Match           11.5%; Score 82; DB 2; Length 530;
Best Local Similarity 25.9%; Pred. No. 23;
Matches  35; Conservative  26; Mismatches  34; Indels  40; Gaps  8;

Qy    9 DYPVEIHESLTALESSLGAVDDMLKTMAVSRNELLQKLDPLEQAKVDLVSAY------ 61
Db 1769 DQKKAVAEQAKLQQQKVNWIQKEIKTNKALNSAQRAQLQEELKQAKLDLISVQDQVRELQ 1828

Qy   62 ---------TLNSMFWVYLATQG-----VNPKEHPVKQELERIRVYMNRVKEITDKKKAA 107
Db 1829 KQLVQSKVDEFLKSIEKSSSKTQGKIKDVDNKISMTEBEDEDKVKYSKQIKLQQQKEA 1888

Qy  108 KLDRGAASRFVKKALWEPKR--KSTPKV 133
Db 1889 K-------KYIKQ-LEEQKAAKGFPDI 1908
```

RESULT 9
S70534
bbK2.10 protein precursor - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S70534
R;Akins, D.R.; Porcella, S.F.; Popova, T.G.; Shevchenko, D.; Baker, S.I.; Li, M.; Norgard,
Mol. Microbiol. 18, 507-520, 1995
A;Title: Evidence for in vivo but not in vitro expression of a Borrelia burgdorferi oute
A;Reference number: S70531; MUID:96342380; PMID:8748034
A;Accession: S70534
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-207 <AKI>
A;Cross-references: UNIPROT:Q44732; EMBL:U18292; NID:g3309519; PIDN:AAC26095.1; PID:g896
C;Genetics:
A;Gene: bbK2.10
C;Superfamily: outer surface protein F ospF
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-207/Product: bbK2.10 protein #status predicted <MAT>

```
Query Match           11.5%; Score 82; DB 2; Length 207;

Qy   10 YPVEIHESLTALESSLGAV-----------DDMLKTMMAVSR----NELLQKLDPLEQAKVD 56
Db  401 YP-KIYQEITVFGKSLVALKPSPSGHAHEDLTKEYLSITEKALFNHLSKSLDPSTFSQVR 459

Qy   57 LVSAYTLNSM--------FWVYLATQGSVNPKEHPVKQELERIRVYMNRVKEITDKKKAAKL 109
Db  460 IM----LNSMNNEDRIFAMFLFT---SAKPEEAKQLLE--KVHVNELKKTID----- 503

Qy  110 DRGAASRFVKKALWE 124
Db  504 ---AFYKFYRDIPWE 515
```

RESULT 10
S70533
bbk2.10 protein precursor - Lyme disease spirochete
C;Species: Borrelia burgdorferi (lyme disease spirochete)
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
A;Accession: S70533
R;Akins, D.R.; Porcella, S.F.; Popova, T.G.; Shevchenko, D.; Baker, S.I.; Li, M.; Norga...
Mol. Microbiol. 18, 507-520, 1995
A;Title: Evidence for in vivo but not in vitro expression of a Borrelia burgdorferi oute...
A;Reference number: S70531; MUID:96342380; PMID:8748034
A;Accession: S70533
A;Molecule type: DNA
A;Residues: 1-207 <AKI>
A;Cross-references: UNIPROT:O44733; EMBL:U19105; NID:g896039; PIDN:AAB40878.1; PID:g8960...
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1994
C;Superfamily: outer surface protein F ospF
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-207/Product: bbk2.10 protein #status predicted <MAT>
C;Genetics:
A;Gene: bbk2.10

Query Match   11.3%; Score 81; DB 2; Length 207;
Best Local Similarity 25.2%; Pred. No. 9.5;
Matches 37; Conservative 25; Mismatches 53; Indels 32; Gaps 5;

```
Qy    2 AGEEMNEDYPVEIHESLITALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSAY 61
          :||::: :::  ||:  ||::|:|:|| |:|||:|:::|::|||:||||||||||||||||
Db   25 SGEDIKQNVKEKVEGFLEAKKEELVGGLKKLGSEAYLKVEELMQADRPQVQAEEQV---- 80

Qy   62 TLINSMFWYLATQGV--NP--KEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRF 117
          ::::  ::|:|     :||:::|| |:|:  |||| |||||||||||||||||||||||||
Db   81 --------AQGVFEDPELKEKGLEEKIEEL---------KELKDSSKKTKEDRKKELEE 122

Qy  118 VKKALWEPKR------KSTPKVANKGK 138
          ::|||:|||||     ::|:||||||||
Db  123 AKQKLEEFKRQVESVTENTDKVKNQGK 149
```

RESULT 11
C59095
hypothetical protein pXO1-35 - Bacillus anthracis virulence plasmid pXO1
C;Species: Bacillus anthracis
C;Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 09-Jul-2004
R;Okinaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler...
J. Bacteriol. 181, 6509-6515, 1999
A;Title: Sequence and organization of pXO1, the large Bacillus anthracis plasmid harbori...
A;Reference number: A59091; MUID:99445483; PMID:10515943
A;Accession: C59095
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-478 <OKI>
A;Cross-references: UNIPROT:Q9X306; GB:AF065404; NID:g4894216; PIDN:AAD32339.1; PID:g489...
A;Note: similar to transposase for insertion sequence element IS231 (478 aa), Bacillus t...

A;Accession: G59095
A;Status: preliminary
A;Molecule type: DNA
A;Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32343.1; PID:g4894255
A;Note: similar to transposase for insertion sequence element IS231 (478 aa), Bacillus t
A;Gene: pXO1-35
A;Genome: plasmid
C;Superfamily: transposase IS231

Query Match   11.3%; Score 81; DB 2; Length 207;
Best Local Similarity 25.2%; Pred. No. 9.5;
Matches 37; Conservative 25; Mismatches 53; Indels 32; Gaps 5;

```
Qy    2 AGEEMNEDYPVEIHESLITALESSLGAVDDMLKTMMAVSRNELLQKLDPLEQAKVDLVSAY 61
          :||::: :::  ||:  ||::|:|:|| |:|||:|:::|::|||:||||||||||||||||
Db   25 SGEDIKQNVKEKVEGFLEAKKEELVGGLKKLGSEAYLKVEELMQADRPQVQAEEQV---- 80

Qy   62 TLINSMFWYLATQGV--NP--KEHPVKQELERIRVYMNRVKEITDKKKAAKLDRGAASRF 117
          ::::  ::|:|     :||:::|| |:|:  |||| |||||||||||||||||||||||||
Db   81 --------AQGVFEDPELKEKGLEEKIEEL---------KELKDSSKKTKEDRKKELEE 122

Qy  118 VKKALWEPKR------KSTPKVANKGK 138
          ::|||:|||||     ::|:||||||||
Db  123 AKQKLEEFKRQVESVTENTDKVKNQGK 149
```

A;Accession: G59095
A;Status: preliminary
A;Molecule type: DNA
A;Cross-references: 77-398,'Q',400-401 <OK2>
A;Experimental source: GB:AF065404; NID:g4894216; PIDN:AAD32343.1; PID:g4894255
A;Note: similar to transposase for insertion sequence element IS231
A;Gene: pXO1-35; pXO1-39
A;Genome: plasmid
C;Superfamily: transposase IS231

RESULT 12
T22976
hypothetical protein F59A2.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
R;Lightning, J.
submitted to the EMBL Data Library, June 1994
A;Reference number: Z19645
A;Accession: T22976
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1133 <WIL>
A;Cross-references: UNIPROT:Q21022; EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F5...
A;Experimental source: clone F59A2
R;Burton, J.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19700
A;Accession: T23157
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1133 <WI2>
A;Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6
A;Experimental source: clone K01A11
C;Genetics:
A;Gene: CESP:F59A2.6
A;Map position: 3
A;Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 635/3; 710/3; 738/3; 795/2; 1008/3; 1...

Query Match   11.3%; Score 81; DB 2; Length 478;
Best Local Similarity 30.2%; Pred. No. 25;
Matches 35; Conservative 13; Mismatches 46; Indels 22; Gaps 5;

```
Qy   44 LQKLDPLEQAKVDLVSAYTLINSMFWV------YLATQGVNPKEHPVKQELERIRVYMN-- 95
          |::|| :::  ::|::::  :::||||      ||||||:||||||||||||||||||||
Db  204 LDDLDQMDRGVTYYISRLKLNNMVVIKNEPPEYFRNGTVKKQSQYIKVDLEHI---MNTL 260

Qy   96 --RVKEITDK---KKAAKLDRGAASRFVKKALWEPKRKSTPKVANKG------KSK 140
           ||||||||   |||||||||||||||||||||||||||||||||       |||
Db  261 EPGQVYEITDAYIGKDKKLFTRVIIYRLTEKQLRERKKQVYTESKKGITYSBKSK 316
```

RESULT 13
T05528

Query Match   11.3%; Score 80.5; DB 2; Length 1133;
Best Local Similarity 26.1%; Pred. No. 73;
Matches 42; Conservative 23; Mismatches 53; Indels 43; Gaps 7;

```
Qy    1 MAGEEMN--EDYPVEIHESLITALESSLGAVDDMLKTMMAV------------SRNELL-- 44
          :|||| :  ||  ||: |: ||::::|: ||||:::|: ||            |||||
Db  361 LAGELENAKEDLKVEEEKHTGIQRAQGALDDAEKEVRVLKEQLERAQSALESSQELASS 420

Qy   45 QKLDPLEQAKVDLVSAY-------TLINSMFWVYLAT-QGVNPKEHPVKQELERIRVYMN 95
          || :||:::  :|:::         ||||||||||| |||||||||||||||||||||||
Db  421 QKADKIQELEKELQNAQKRSSEELETANEMVRSLTATLENSNSETEILKQKLETL---- 475

Qy   96 RVKEITDKGRGAASRFVKKALWEPKRKSTPKVANK 136
          :: :::            ::||||||||||||||||
Db  476 ------DKELQAR------QQTEKALTEIRINVLTTSIAEK 503
```

hypothetical protein F13M23.210 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T05528
R;Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15419
A;Accession: T05528
A;Molecule type: DNA
A;Residues: 1-317 <BEV>
A;Cross-references: UNIPROT:Q9SW19; EMBL:AL035523
A;Experimental source: cultivar Columbia; BAC clone F13M23
C;Genetics:
A;Map position: 4
A;Note: F13M23.210
A;Introns: 23/3; 48/3; 98/3; 119/1; 157/1; 196/3; 220/1; 233/3; 276/1

Query Match   11.2%; Score 80; DB 2; Length 317;
Best Local Similarity 24.2%;
Matches   31; Conservative 27; Mismatches 46; Indels 24; Gaps   6;

Qy   16 ESLTALESSLGAVDDMLI----KTMMAVSR-----------NELLQKQLDPLEQAKVDLVSAYTLNS 65
         |::|||:|||| |||||:: ||| ||| :||||||||||||||||||||:|:|::
Db 180 ESMLAVETGLREIASLKVEDAVMLALSRYRQTNVARQAVTELSHBQQDILFKEAWLL-- 237

Qy   66 MFWVYLATQGVNPKEHPVKQEL--ERIRVVMNRVKEITDKKAAKLDRG---AASRFVKK 120
          :|||||||:||||:|||||||: |||:|||||||:||||||||||:: |||| ||::
Db 238 YFW------KRAKIHGVESDIAEERLQFWINRLGQHSSSHDAIDVERGMRELRKLGIEQ 290

Qy 121 ALWEPKRK 128
         ||:|::||
Db 291 QLWETSRK 298

RESULT 14
S54396
protein L precursor - Peptostreptococcus magnus
C;Species: Peptostreptococcus magnus
A;Variety: strain 3316
A;Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S54396
R;Murphy, J.P.; Duggleby, C.J.; Atkinson, M.A.; Trowern, A.R.; Atkinson, T.; Goward, C.R
Mol. Microbiol. 12; 911-920; 1994
A;Title: The functional units of a peptostreptococcal protein L.
A;Reference number: S54396; MUID:95020613; PMID:7934898
A;Accession: S54396
A;Molecule type: DNA
A;Status: preliminary; nucleic acid sequence not shown
A;Residues: 1-992 <MUR>
A;Cross-references: UNIPROT:Q51918; EMBL:L04466; NID:g150673; PIDN:AAA67503.1; PID:g1506

Query Match   11.2%; Score 80; DB 2; Length 992;
Best Local Similarity 23.6%;
Matches   29; Conservative 32; Mismatches 42; Indels 20; Gaps   5;

Qy   9 DYPVEIHESLTALESSLGAVDDMLKTMMAVSRNELLQKQLDPLEQ-------AKVDLVSA 60
          |::| ||||:||||||||| :||||||:::|||||||:::||| |||||||
Db   47 DYHGDVSDSVDPVEEE---IDEALAKALAEAKETAKKHIDSLNHLSETAKKLAKNDIDSA 103

Qy   61 YTLNSMFWVYLATQGVNPKEHPVKQELERI------RVVMNRVKEITDK-KKAAKLDRG 112
         |||:|::|||||||::|||||||||||||| |||||||||||||: ||||||||
Db 104 TTINAINDI-VARADVMERKTAEKEEAEKLAAAKETAKKHIDELKHLADKTKELAKRDID 162

Qy 113 AAS 115
         ||:
Db 163 SAT 165

RESULT 15
E64542
hypothetical protein HP0181 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004

C;Accession: E64542
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
C;Date: 07-Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watchey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: E64542
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-224 <TOM>
A;Cross-references: UNIPROT:O24983; GB:AE000538; GB:AE000511; NID:g2313263; PIDN:AAD0725:
C;Genetics:
A;Start codon: GTG

Query Match   11.0%; Score 79; DB 2; Length 224;
Best Local Similarity 24.8%;
Matches   34; Conservative 27; Mismatches 54; Indels 22; Gaps   6;

Qy   12 VEIHESLTALESSLGAVDDMLKTMM-------AVSRNELLQKQLDPLEQAKVDLVSAY--T 62
         |:|||||:|||||||| |:||||: |:|||:||||||||:|:|||||||| |
Db   82 VLVFSGLGIIDKALGFIFSCLKTFLVLSFILYALSKMEVMKDANAYLQEK---SAFFST 137

Qy   63 LNSMFWVYLATQGVNPKEHPVKQELERIRVVMNRVKEITDKKKAA-----KLDRGAAS--- 115
         ||:||||||||:|||||||||||||::|||||||||||||||: |||||||||:
Db 138 MRSVASKIMRLDGVKHVEQNIKDNLEEMSDEVKN-KESFNQKESFNKAMDKGVESLKEK 196

Qy 116 -RFVKKALWEPKRKSTP 131
          |||| ||:||:||:||
Db 197 AKDLPKQMLDPKANQTP 213

Search completed: January 4, 2005, 13:45:15
Job time : 18.5 sec

This Page Blank (uspto)

OM protein - nucleic search, using frame_plus_p2n model

```
                              GenCore version 5.1.6
                              Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:                       January 5, 2005, 13:12:38

Title:                        US-09-701-618A-4
Perfect score:                715
Sequence:                     1 MAGEMNEDYPVEIHESLTA..........LWEPKRKSTPKVANKGKSKH 141

Scoring table:                BLOSUM62
                              Xgapop   10.0 , Xgapext   0.5
                              Ygapop   10.0 , Ygapext   0.5
                              Fgapop    6.0 , Fgapext   7.0
                              Delop     6.0 , Delext    7.0

Searched:                     824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters:        1649014

Minimum DB seq length:        0
Maximum DB seq length:        2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Command line parameters:
```
-MODEL=frame+_p2n.model  -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09701618/runat_05012005_085839_24463/app_query.fasta_1.654
-DB=Issued_Patents_NA  -QFMT=fastap  -SUFFIX=rni  -MINMATCH=0.1  -LOOPCL=0
-LIST=45  -DOCALIGN=200  -UNITS=bits  -START=1  -END=-1  -MATRIX=blosum62  -TRANS=human40.cdi
-LOOPEXT=0  -THR_SCORE=pct  -THR_MAX=100  -THR_MIN=0  -ALIGN=1
-MODE=LOCAL  -OUTFMT=pto  -NORM=ext  -HEAPSIZE=500  -MINLEN=0  -MAXLEN=200000000
-USER=US09701618 @CGN 1 1 141 @runat 05012005 085839 24463  -NCPU=6  -ICPU=3
-NO_MMAP  -LARGEQUERY  -NEG_SCORES=0  -WAIT  -DSPBLOCK=100  -LONGLOG
-DEV TIMEOUT=120  -WARN_TIMEOUT=30  -THREADS=1  -XGAPOP=10  -XGAPEXT=0.5  -FGAPOP=6
-FGAPEXT=7  -YGAPOP=10  -YGAPEXT=0.5  -DELOP=6  -DELEXT=7
```

Database :   Issued_Patents_NA.*
```
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq;*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq;*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq;*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq;*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq;*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq;*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	485	67.8	380	4	US-09-513-999C-1235	Sequence 1235, Ap
2	113	15.8	651	4	US-09-248-796A-841	Sequence 841, App
3	85.5	12.0	885	4	US-09-134-000C-332	Sequence 332, App
4	84	11.7	819	4	US-09-248-796A-4954	Sequence 4954, Ap
5	82.5	11.5	7100	4	US-09-308-375-1	Sequence 1, Appli
6	82	11.5	924	4	US-09-107-532A-3386	Sequence 3386, Ap
7	80.5	11.3	1407	4	US-09-248-796A-13211	Sequence 13211, A
8	80.5	11.3	1428	4	US-09-248-796A-1115	Sequence 1115, Ap
9	80.5	11.3	1710	3	US-08-747-221B-33	Sequence 33, Appl
10	80.5	11.3	1710	4	US-09-005-051-33	Sequence 33, Appl
11	80.5	11.3	1710	4	US-09-403-942F-33	Sequence 33, Appl
12	80.5	11.3	1785	3	US-08-747-221B-34	Sequence 34, Appl
13	80.5	11.3	1785	3	US-08-747-221B-35	Sequence 35, Appl
14	80.5	11.3	1785	4	US-09-005-051-34	Sequence 34, Appl
15	80.5	11.3	1785	4	US-09-005-051-35	Sequence 35, Appl
16	80.5	11.3	1785	4	US-09-403-942F-34	Sequence 34, Appl
17	80.5	11.3	1785	4	US-09-403-942F-35	Sequence 35, Appl
18	80.5	11.3	2801	3	US-08-747-221B-30	Sequence 30, Appl
19	80.5	11.3	2801	3	US-08-747-221B-32	Sequence 32, Appl
20	80.5	11.3	2801	4	US-09-005-051-30	Sequence 30, Appl
21	80.5	11.3	2801	4	US-09-005-051-32	Sequence 32, Appl
22	80.5	11.3	2801	4	US-09-403-942F-30	Sequence 30, Appl
23	80.5	11.3	2801	4	US-09-403-942F-32	Sequence 32, Appl
24	80.5	11.3	1664976	4	US-09-403-942F-32	Sequence 32, Appl
25	80	11.2	3279	4	US-08-446-137B-1	Sequence 1, Appli
26	79	11.0	1664976	4	US-08-916-421B-1	Sequence 1, Appli
27	79	11.0	1230025	4	US-08-198-452A-1	Sequence 1, Appli
28	78.5	11.0	1710	3	US-08-747-221B-27	Sequence 27, Appl
29	78.5	11.0	1710	4	US-09-005-051-27	Sequence 27, Appl
30	78.5	11.0	1710	4	US-09-403-942F-27	Sequence 27, Appl
31	78.5	11.0	1788	3	US-08-747-221B-28	Sequence 28, Appl
32	78.5	11.0	1788	4	US-09-005-051-28	Sequence 28, Appl
33	78.5	11.0	1788	4	US-09-403-942F-28	Sequence 28, Appl
34	78.5	11.0	1788	3	US-09-005-051-29	Sequence 29, Appl
35	78.5	11.0	1788	4	US-09-005-051-29	Sequence 29, Appl
36	78.5	11.0	1788	4	US-09-403-942F-29	Sequence 29, Appl
37	78.5	11.0	2836	3	US-08-747-221B-24	Sequence 24, Appl
38	78.5	11.0	2836	3	US-08-747-221B-26	Sequence 26, Appl
39	78.5	11.0	2836	4	US-09-005-051-24	Sequence 24, Appl
40	78.5	11.0	2836	4	US-09-005-051-26	Sequence 26, Appl
41	78.5	11.0	2836	4	US-09-403-942F-24	Sequence 24, Appl
42	78.5	11.0	2836	4	US-09-403-942F-26	Sequence 26, Appl
43	78	10.9	456	4	US-09-134-000C-3074	Sequence 3074, Ap
44	77.5	10.8	579	4	US-09-134-000C-3226	Sequence 3226, Ap
45	77	10.8	2433	4	US-09-248-796A-6178	Sequence 6178, Ap
```

ALIGNMENTS

RESULT 1
US-09-513-999C-1235

; Sequence 1235, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent Reference: 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1235
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 75..380
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9
; OTHER INFORMATION: a=g or c
US-09-513-999C-1235

Alignment Scores:
Pred. No.:   9.92e-62
Score: 485.00
Percent Similarity: 98.04%
Best Local Similarity: 92.16%
Query Match: 67.83%

Length: 380
Matches: 94
Conservative: 6
Mismatches: 2
Indels: 0

```
                                          4                      Gaps:             0
```

```
DB:

Qy   US-09-701-618A-4 (1-141) × US-09-513-999C-1235 (1-380)

Db   US-09-513-999C-1235

Qy     1  MetAlaGlyGluGlyMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAla  20
Db    75  ATGGCAGGTGGAGAAATTAATGAAGACTATCCAGTAGAAATTCACGAGTATTTGTCAGCG  134

Qy    21  LeuGluSerLeuGlyAlaValAlaAspAspMetLeuLysThrMetAlaValSerArg  40
Db   135  TTTGGAATTCAACATGGTGTCTGGATGAGAAGTGAAGACCATGATGTCTGTTTCTAGA  194

Qy    41  AsnGluLeuLeuGlnLysLeuAspProLeuGlnIleAlaLysValAspLeuVeuValSerAla  60
Db   195  AATGAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCA  254

Qy    61  TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnInGlyValAsnProLysGlu  80
Db   255  TACACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGAGTTAATCCTAAGGAA  314

Qy    81  HisProValLysGlnGluLeuArgIleArgValTyrMetAsnArgValLysGluIle  100
Db   315  CATCCAGTAAAACAGGAATTGAGAATTCAAGAGTATATGAACAGAGTCAAGGAAATA  374

Qy   101  ThrAsp  102
Db   375  ACAGAC  380
```

RESULT 2

US-09-248-796A-841

; Sequence 841, Application US/0924879 6A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 841
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-841

Alignment Scores:
Pred. No.:   4
Score: 113.00
Percent Similarity: 44.17%
Best Local Similarity: 27.50%
Query Match: 15.80%

Length: 651
Matches: 33
Conservative: 20
Mismatches: 47
Indels: 20
Gaps: 2

US-09-701-618A-4 (1-141) × US-09-248-796A-841 (1-651)

```
Qy    10  TyrProValGluIleHisGluSerLeuThrAlaLeuGluSerLeuGly------  26
Db   199  TATAAAACTCATTATCACAATCAATATCAGCTTATGAATTCCATTAGTTCATTACAA  254
```

RESULT 3

US-09-134-000C-332

; Sequence 332, Application US/0913400 0C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-15
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 332
; LENGTH: 885
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-332

Alignment Scores:
Pred. No.:   4
Score: 0.0126
Percent Similarity: 85.50
Best Local Similarity: 42.76%
Query Match: 11.96%

Length: 885
Matches: 38
Conservative: 24
Mismatches: 50
Indels: 33
Gaps: 6

US-09-701-618A-4 (1-141) × US-09-134-000C-332 (1-885)

```
Qy     1  MetAlaGlyGluGlyMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAla  20
Db   199  ATGGCTGGGTGAAGGTTTTGCTCGAAGATCACCAAGAAGAGTTAGAACAGGTTTTAACC--  255

Qy    21  LeuGluSerLeuGlyAlaValAlaAspAspMetLeuLysThrMetAlaValSerArg  40
Db   256  --------------ATGAATGAGTGCCTTCATTCAAGAGAATTTTAGTGAACTGGGAAAAT  294

Qy    41  AsnGluLeuLeuGlnLysLeuAspProLeuGlnIleAlaLysValAspLeuVeuValSerAla  60
Db   295  CAAGAGCTGTATCAAACAAATACCCTGTTGAAGATCTTACAGATTAGTTCGGAAAAGAGC  348
```

RESULT 4
US-09-248-796A-4954
Sequence 4954, Application US/09248796A
Patent No. 6747137

; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4954
; LENGTH: 819
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (41),(249)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno

US-09-248-796A-4954

Alignment Scores:
Pred. No.:
Score: 0.0185
Percent Similarity: 84.00
Best Local Similarity: 41.83%
Query Match: 21.57%
DB: 11.75%

Length: 819
Matches: 33
Conservative: 31
Mismatches: 51
Indels: 38
Gaps: 5

US-09-701-618A-4 (1-141) x US-09-248-796A-4954 (1-819)

```
Qy      4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer  23
            |||              :::                 ::||||||||:::
Db    148 GAATCATATAGTGAAGATAATCAAGCAATCCATGTCGAAACTATCACTTCTTGAAACA  207

Qy     24 SerLeuGlyAlaValAspAsp---MetLeuLysThrMetMetAlaValSerArgAsnGlu  42
            :::|||    ::     :::||||    :::  ||| :::
Db    208 GCATTGTCAGAGAAAAATGAAAAATTATTAGAAACACAGAGAAGAGAACTGAA  267

Qy     43 LeuLeuGlnLysLeuAspProLeuGlnIleAlaLysValAspLeuVal------------  58
            :::|||||        :::    |||
Db    268 -----GAGGCATTAGACGCCTGAATTACAATAAGGAAAACCGATTTACCAAAATTAAGTGAC  321

Qy     59 ------SerIleTyrThrLeuAsnSerMetPheTrpVal  69
                ||||:::||||:::::|||||||
Db    322 TCCTCGGCTATAGATAGCTGCTAATGGTAGTGGTTATTCTTTATCAACGTTATATTCAGAG  381

Qy     70 Tyr-----LeuAlaThrGlnGlyValAsnProLysGluHisProValLysGlnGluLeu  87
            |||       :::    |||     :::|||
Db    382 TATAACCATTTGAAGAAACAATTCGGTGGTAGAAAAGATCTCAAAAACAAAAAAATGGAAATG  441

Qy     87 -------------------------------------  87
Db    442 CAACTTGAGTCATTATTAATTGAATTGGTAGCTAGAAAGCCAGTTATTGCCAATTACCGT  501

Qy     88 GluArgIleArgValTyrMetAsnArgValLysGluIleThrAspLysGlyLysAlaAla  107
            :::|||:::|||:::||||||    |||:::    :::
Db    502 GAACAAATTCAATTTATGAAAAATTCCATGAAAGAGAGATGCTTGGTTGTAAAGTTGAAACCATA  561

Qy    108 LysLeuAspArgGlyTyrAlaAlaSerArgPheValLysLys  120
            :::|||:::    |||
Db    562 AGATCAGAAAAAGGTGAAGTTGAAAAAGAATGCCAAGAGA  600
```

RESULT 5
US-09-308-375-1
; Sequence 1, Application US/0930375

US-09-308-375-1

Alignment Scores:
Pred. No.:
Score: 0.972
Percent Similarity: 82.50
Best Local Similarity: 43.24%
Query Match: 22.30%
DB: 11.54%

Length: 7100
Matches: 33
Conservative: 31
Mismatches: 53
Indels: 31
Gaps: 5

US-09-701-618A-4 (1-141) x US-09-308-375-1 (1-7100)

```
Qy      9 AspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSerSerLeuGlyAlaVal  28
            ::: |||              :::  ::||||||||:::||  ::
Db   5416 GATCAGAAAAAAGCTGTGGCAGCGAAGCTAAAATCCAACAACAAAAAAGTTAATTGGATT  5475

Qy     29 AspAspMetLeuLysThrMetMetAlaValSerArgAsnGluLeuLeuGlnLysLeuAsp  48
            :::||| :::    ||| :::    :::    |||  :::
Db   5476 CAAAAAGAAATTAAAACAAATAAAAGCATTAGAACTCCGCTCAAGCTGCACAGCTTCAAGAA  5535

Qy     49 ProLeuGlnIleAlaLysValAspLeuValSerAlaTyr------------------  61
            |||     :::         :::
Db   5536 GAGCTTAAACAGGCCAAGCTAGATTTAATTCGTTCAAGACCAGGTTCGTGAGCTACAG  5595

Qy     62 ------------------------------ThrLeuAsnSerMetPheTrpValTyrLeu  71
                          ||||:::||||:::
Db   5596 AAACAACTTGTTCAATCTAAAGTTGATGAGACACAGAAGCTAAGTGCAATTGAAAAGTCATCTTCT  5655

Qy     72 AlaThrGlnGly------ValAsnProLysGluHisProValLysGlnGluLeu  87
            ||||||| |||        :::  |||  :::|||
Db   5656 AAAACCCAAGGGAAAAATTAAAGATGTCGATAACAAAAATTTCAATGACGTGAAGATGAA  5715

Qy     88 GluArgIleArgValTyrMetAsnArgValLysGluIleThrAspLysGlyLysAlaAla  107
            :::|||:::||||:::||||||    |||:::    :::
Db   5716 GACAAGGTTAAATACTATTAGCGAAGCAAATAAAGCTCATTCAACACACAAAAAAGGGAGCG  5775

Qy    108 LysLeuAspArgGlyTyrAlaAlaSerArgPheValLysLysAlaLeuGluThrProLysArg  127
            :::|||:::    |||                       |||:::
Db   5776 AAG-------AAATACATTAAGCAG---CTTGAAGAACAAAAGGAAA  5811

Qy    128 -----LysSerThrProLysVal  133
                  |||  :::
Db   5812 GCTGCGAAAAGGTTTCCCTGACATC  5835
```

RESULT 6
US-09-107-532A-3386
; Sequence 3386, Application US/09107532A
; Patent No. 6583275

```
ZIP: 02354
COMPUTER READABLE FORM:
    MEDIUM TYPE: CD/ROM ISO9660
    COMPUTER: PC
    OPERATING SYSTEM: <Unknown>
    SOFTWARE: ASCII
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/107,532A
    FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 60/085,598
    FILING DATE: 14 May 1998
    APPLICATION NUMBER: 60/051571
    FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
    NAME: Ariniello, Pamela Deneke
    REGISTRATION NUMBER: 40,489
    REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (781)893-5007
    TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3386:
    SEQUENCE CHARACTERISTICS:
    LENGTH: 924 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: circular
    MOLECULE TYPE: DNA (genomic)
    HYPOTHETICAL: NO
    ANTI-SENSE: NO
    ORIGINAL SOURCE:
    ORGANISM: Enterococcus faecium
    FEATURE:
        NAME/KEY: misc_feature
        LOCATION: (B) LOCATION 1...924
    SEQUENCE DESCRIPTION: SEQ ID NO: 3386:
US-09-107-532A-3386

Alignment Scores:
Pred. No.:
Score:                       0.044    Length:         924
                                      Matches:         32
Percent Similarity:         49.26%    Conservative:    35
Best Local Similarity:      23.53%    Mismatches:      33
Query Match:                11.47%    Indels:          36
DB:                          4        Gaps:             8

US-09-701-618A-4 (1-141) x US-09-107-532A-3386 (1-924)

Qy    3 GlyGluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGlu 22
        |||:|::|::|:::|
Db  235 GGGCAAGATGTATGGGGAAGAT---AGTTTATCTATCCATGAAACGTTCTTCTTATNGTTCCA 291

Qy   23 SerSerLeu-----GlyAlaValAspAspMetLeuLysThr 34
        ||:|
Db  292 GGCGATATCTCTTCTTCGTGGGGAAAATTGACTGGGAGGACGGATCATCGATCTTCATGAAG 351

Qy   35 Met---------MetAlaValSerArgAsnGlyIleLeuGlnLys-----LeuAsp 48
        |||
Db  352 CTTCATGCAGGAAAAGGAGAGTCGCTCAAAAAAGATGAACTCGATTCAACGTTTCTCTCTGAT 411

Qy   49 ProLeuGluGlnAla--------LysValAspLeuValSer 59
        |||::|
Db  412 CCTAAGAAAAAAGCAAAAAACTATTCAAAAAGGAAAATCGCCAAAAAGTGGGCTTGATTGCT 471

Qy   60 AlaTyrThrLeuAsnSerMetPheTrpValTyr--LeuAlaThrGlnGlyValAsnPro 78
        ||:
Db  472 GCTTTGGCTGTTGATTCGGATCTCTATTTATTTGATGAGCCTACTTCTCGGTTTAGATCCT 531

Qy   79 ---LysGluHisProValLysGlnGluLeuGlnArgIleArgVal------- 92
        |||
Db  532 TTGATGGAAACGACATTTTCCAAGAAGAGTTGAAAAATTGAAGAACAGGGCAAATCCATC 591

Qy   93 -------TyrMetAsnArgValLysGluGluIleThrAspLys 103
```

```
Db  592 CTACTTTCTTCTTCATATTCTCACTGAAGTCGACCGACTAGCGCGGATAAA 639
        :::|::::|                     |||:::  ::  |||||||

RESULT 7
US-09-248-796A-13211
Sequence 13211, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
    APPLICANT: Keith Weinstock et al
    TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
    TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
    FILE REFERENCE: 107196.132
    CURRENT APPLICATION NUMBER: US/09/248,796A
    CURRENT FILING DATE: 1999-02-12
    PRIOR APPLICATION NUMBER: US 60/074,725
    PRIOR FILING DATE: 1998-02-13
    PRIOR APPLICATION NUMBER: US 60/096,409
    PRIOR FILING DATE: 1998-08-13
    NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 13211
    LENGTH: 1407
    TYPE: DNA
    ORGANISM: Candida albicans
US-09-248-796A-13211

Alignment Scores:
Pred. No.:
Score:                       0.143    Length:        1407
                                      Matches:         29
Percent Similarity:         80.50     Conservative:    20
Best Local Similarity:      45.79%    Mismatches:      35
Query Match:                27.10%    Indels:          23
DB:                          4        Gaps:             5

US-09-701-618A-4 (1-141) x US-09-248-796A-13211 (1-1407)

Qy    3 GlyGluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGlu 22
        |||:::::::|
Db  793 GGGGACAAGTTGTACGAGAGATGGAGTGGAGTTCAATGAA---TTATTGAAAAATGGAC 849

Qy   23 SerSerLeuGlyAlaValAspAsp---AspMetLeuLysThrMetMetAlaValSerArgAsn 41
        :::||
Db  850 TTAAATCTACCCAAGCAGGATAAGGAAATTACTGCATAGNTATTATGAGGTGCTTACTGCGAC 909

Qy   42 ---------GluLeuGlnLysLeuAspProLeuGlu 51
        |||
Db  910 TTGCAAGAGATGTCGGTAAAGTGCGGAGACATTTTTGTCAATGAGTTGGAACCGAAGGAT 969

Qy   52 GlnAlaLysValAsp--------LeuValSerAlaTyrThrLeuAsnSer 65
        |||:||
Db  970 GCGGGTGGAGTTGATGAAACAAGCGCAACGATAATGATCAAAGCACATATACATAATGAC 1029

Qy   66 Met-----------PheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        :::|
Db 1030 TTGATCTATGTTCAGAAATTTGGGATAACCTCAAGTTTGGATTGGTTCCGGATTAT 1089

Qy   81 HisProValLysGlnGluLeu 87
        |||||||
Db 1090 CAAATGTTCAAGCAATACTTG 1110

RESULT 8
US-09-248-796A-1115
Sequence 1115, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
    APPLICANT: Keith Weinstock et al
    TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
    TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
    FILE REFERENCES: 107196.132
    CURRENT APPLICATION NUMBER: US/09/248,796A
    CURRENT FILING DATE: 1999-02-12
    PRIOR APPLICATION NUMBER: US 60/074,725
    PRIOR FILING DATE: 1998-02-13
    PRIOR APPLICATION NUMBER: US 60/096,409
```

Column 1

```
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 1115
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-1115
DB:
Query Match:
```

Alignment Scores:
```
Pred. No.:
Score:                     0.147          Length:          1428
Percent Similarity:        80.50          Matches:         29
Best Local Similarity:     46.73%         Conservative:    21
Query Match:               27.10%         Mismatches:      38
                           11.26%         Indels:          19
                           4              Gaps:            4
```

US-09-701-618A-4 (1-141) x US-09-248-796A-1115 (1-1428)

```
Qy    4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
          |||  |||   ||:|  |||:   |
Db  418 GAATCCCTTCGGTGAAGTCTACCCCGAAGTTTTGGGGTCAATTTAAATGCTTTATATTCA 477

Qy   24 SerLeuGlyAlaValAsp---AspMetLeuLysThrMetAlaValSerArgAsnGlu 42
          :::|||:||   ::|      :::  ||||||
Db  478 TGTCTTGATTCAATTGACAAAATCAACATTATATACCATTGTCAAAATCCTTCAATTAATCAA 537

Qy   43 LeuLeuGlnLysLeuAspProLeu------GluGlnAlaLys 54
          ||| |||:|:                  :|| |||
Db  538 ATCCTACCAACATTAACACCAATTTTAAAAATCGACACGAAAAAGTTCAAGACGCATGC 597

Qy   55 ValAspLeuVal------------SerAlaTyrThrLeuAsnSerMetPheTrp 68
          ||||  ||                || ::: :|
Db  598 ATAAAATTGGTTGGATTAATATAGCTCTGAAACTCTATAAACCGAAAGAATGG 657

Qy   69 ValTyrLeuAlaThrGlnGlyValAsnProLysGlyGluHisProValLysGlnGluLeuGlu 88
          ||| ::: |||                   ::|
Db  658 ATGGGTATTGTTGTTGATTGTTG------GAAATGTTAAAAGCCCAACGTAAA 705

Qy   89 ArgIleArgValTyrMetAsn 95
          ||||||||||| ::|  |||
Db  706 AGAATACGTATTGCAGCTAAT 726
```

```
RESULT 9
; Sequence 33, Application US/08472221B
; Patent No. 6063610
; GENERAL INFORMATION:
;   APPLICANT: Silver, Gary W.
;   APPLICANT: Wisnewski, Nancy
;   TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
;   NUMBER OF SEQUENCES: 66
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Carol Talkington Verser, Ph.D.
        ADDRESSEE: Heska Corporation
        STREET: 1825 Sharp Point Drive
        CITY: Fort Collins
        STATE: Colorado
        COUNTRY: USA
        ZIP: 80525
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: Windows 95
   SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/747,221B
   FILING DATE: No. 6063610ember 12, 1996
   CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
   NAME: Verser, Carol Talkington
   REGISTRATION NUMBER: 37,459
   REFERENCE/DOCKET NUMBER: FC-1
```

Column 2

```
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 970/493-7272
;   TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 33:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1710 nucleotides
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..1710
US-08-747-221B-33
DB:
```

Alignment Scores:
```
Pred. No.:
Score:                     0.196          Length:          1710
Percent Similarity:        80.50          Matches:         38
Best Local Similarity:     38.41%         Conservative:    25
Query Match:               23.17%         Mismatches:      62
                           11.26%         Indels:          39
                           3              Gaps:            8
```

US-09-701-618A-4 (1-141) x US-08-747-221B-33 (1-1710)

```
Qy    4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
          :::||| |||:||||                      :|
Db 1138 AAAGAACAAAAACGATGACATC--------------GAAGCCTATGTCCAACTAAAAGGC 1182

Qy   24 SerLeuGlyAlaValAspAspMetLeuLysThrMetAlaValSerArgAsnGluLeu 43
          :::  ||||  ||                    ||||||:|||||:|
Db 1183 GATGCTGGTTATCTCCAAGGAATTTACCGTACCTTGAAAGCCATATTTTTCAATGAAATC 1242

Qy   44 LeuGlnLys------------LeuAspProLeuGluGlnAlaLysValAspLeuValSer 59
          |||||                :::
Db 1243 AAAAGAAAATTCCAACTTGTATTTGTTTATAGGTTATCAGATGATACGTATAGTGTATATAAA 1302

Qy   60 AlaTyrThrLeu------AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
          :::|||                |||||| :::
Db 1303 AGTTATATCTTGCCCTATCGATGGGGTTCCTTGCCAGGAGTTAGTCATGGTGATGATTTA 1362

Qy   64 --------------AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
                        ||||||  :::
Db 1423 CCGCAAGAT-----GCTATGCAGACTCTGGAAAGGATGGGTCAGGATCTGGACCAATTTT 1476

Qy   78 ProLysGluHisProValLysGlnGluLeuGluArg---IleArgValTyrMetAsnArg 96
          |||  ::::                 :::
Db 1363 GGATATCTTTTTGCAAACTCTTTGGATGTTGCCTATTTGGGAACAACGACACATTTCTATA 1422

Qy   97 ValLys-----GluIleThrAspLysValSerLysAlaAlaIleLysLeuLeuAspArgGlyAlaAla 114
          ||||||            :::    |||    :::
Db 1477 GTAAAGAATGGAAAAACCTACATCAAACACTGAAGATGCATCATGTGAT------ACA 1527

Qy  115 SerArgPheValLysAlaLysGlnLeuTrpGluProLysAspArgLysSerThrProLysGlnGlyAla 134
          |||||:                    ::||||
Db 1528 AAAAGACATTTAAAGACATTTTTTTGGGAACCATACAACGACGAGAACCAAAAAAAAATATTTG 1587

Qy  135 AsnLysGlyLys 138
          ||||||
Db 1588 GACATCGGAAA 1599
```

```
RESULT 10
US-09-005-051-33
; Sequence 33, Application US/09005051
; Patent No. 6291222
; GENERAL INFORMATION:
;   APPLICANT: Silver, Gary W.
;   APPLICANT: Wisnewski, Nancy
;   TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
;   NUMBER OF SEQUENCES: 66
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Carol Talkington Verser, Ph.D.
```

ADDRESSEE: Heska Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: November 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: PC-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 1710 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1710

US-09-005-051-33

Alignment Scores:
Pred. No.:
Score: 0.196
Percent Similarity: 80.50
Best Local Similarity: 38.41%
Query Match: 11.26%
DB:

Length: 1710
Matches: 38
Conservative: 25
Mismatches: 62
Indels: 39
Gaps: 8

US-09-701-618A-4 (1-141) x US-09-005-051-33 (1-1710)

Qy   4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
         ::||:|| ||||||:||| |:||
Db 1138 AAAGAACAAAACGATGACATC--------GAAGCCTAGTCCAACTAAAAGGC 1182

Qy   24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValSerArgAsnGluLeu 43
         |||||||| || ::: :|||||||:|:
Db 1183 GATGCTGGTTATCTCCAAGGAAGTTACCGTACCTTGAAAGCCATATTTTCAATGAAATC 1242

Qy   44 LeuGlnLys------------LeuAspProLeuGlnIleAlaLysValAspLeuValSer 59
         ::: |||
Db 1243 AAAAGAAATTCCAACTTGTATTTGTATAGGTTATCAGATGATAACGTATATAGTGTATAAA 1302

Qy   60 AlaTyrThrLeu 63
            |||
Db 1303 AGTTATATCTTGCCCTATCGATGGGTTCCTTGCCAGGAGTTAGTCATGGTGATGATTA 1362

Qy   64 ----------AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
                     :::::|| |||
Db 1363 GGATATCTTTTGCAAACTCTTTGGATGTTCCTATTTTGGGAACAACGCACATTTCTATA 1422

Qy   78 ProLysGluHisProValLysGlnGluGluArg---IleArgValTyrMetAsnArg 96
         ||| |:|||::: |||
Db 1423 CCGCAAGAT-----GCTATGCAGACTCTGGAAAAGATGGGTCAGGATCTGACCAATTTT 1476

Qy   97 ValLys----------GluIleThrAspLysLysGlyLysAlaAlaLysLeuAspArgGlyAlaAla 114
         |||||| ::: |||
Db 1477 GTAAAGAATGGAAAACCTACATCAAACACTGAAGAGTGCATCATGTGAT--------ACA 1527

Qy 115 SerArgPheValLysLysAlaLeuTrpGluProLysArgLysSerThrProLysValAla 134
         :::||| |||||| || ||||||
Db 1528 AAAAGACAATTAAACGACATTTTTTGGGAACCATACAACGACGAAGAACCAAAAATATTG 1587

Qy 135 AsnLysGlyLys 138
         ::: ||||||
Db 1588 GACATGGGAAAA 1599

RESULT 11
US-09-403-942F-33
; Sequence 33, Application US/09403942F
; Patent No. 6664090
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary M.
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Wisnewski, Nancy
; TITLE OF INVENTION: NOVEL CARBOXYLESTERASE NUCLEIC ACID MOLECULES, PROTEINS AND USES
; FILE REFERENCE: PC-1-C1-PUS
; CURRENT APPLICATION NUMBER: US/09/403,942F
; CURRENT FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: PCT/US97/20598
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 08/747,221
; PRIOR FILING DATE: 1996-11-12
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 1710
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(1710)
; OTHER INFORMATION:
US-09-403-942F-33

Alignment Scores:
Pred. No.:
Score: 0.196
Percent Similarity: 80.50
Best Local Similarity: 38.41%
Query Match: 11.26%
DB:

Length: 1710
Matches: 38
Conservative: 25
Mismatches: 62
Indels: 39
Gaps: 8

US-09-701-618A-4 (1-141) x US-09-403-942F-33 (1-1710)

Qy   4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
         ::||:|| ||||||:||| |:||
Db 1138 AAAGAACAAAACGATGACATC--------GAAGCCTAGTCCAACTAAAAGGC 1182

Qy   24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValSerArgAsnGluLeu 43
         |||||||| || ::: :|||||||:|:
Db 1183 GATGCTGGTTATCTCCAAGGAAGTTACCGTACCTTGAAAGCCATATTTTCAATGAAATC 1242

Qy   44 LeuGlnLys------------LeuAspProLeuGlnIleAlaLysValAspLeuValSer 59
         ::: |||
Db 1243 AAAAGAAATTCCAACTTGTATTTGTATAGGTTATCAGATGATAACGTATATAGTGTATAAA 1302

Qy   60 AlaTyrThrLeu 63
            |||
Db 1303 AGTTATATCTTGCCCTATCGATGGGTTCCTTGCCAGGAGTTAGTCATGGTGATGATTA 1362

Qy   64 ----------AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
                     :::::|| |||
Db 1363 GGATATCTTTTGCAAACTCTTTGGATGTTCCTATTTTGGGAACAACGCACATTTCTATA 1422

Qy   78 ProLysGluHisProValLysGlnGluGluArg---IleArgValTyrMetAsnArg 96
         ||| |:|||::: |||
Db 1423 CCGCAAGAT-----GCTATGCAGACTCTGGAAAAGATGGGTCAGGATCTGACCAATTTT 1476

Qy   97 ValLys----------GluIleThrAspLysLysGlyLysAlaAlaLysLeuAspArgGlyAlaAla 114
         |||||| ::: |||

Db   1477 GTAAAGAATGGAAAACCTACATCAAACACTGAAGATGCATCATGTGAT--------ACA 1527
           ||| ::: |||
Qy   115 SerArgPheValLysLysAlaLeuThrpGluProLysSerThrProLysValAla 134

Db   1528 AAAAGACATTTAAACGACATTTTTGGGAACCATACAACGACGAGAACCAAAATATTTG 1587
           :::: ||||| |||||
Qy   1588 GACATGGGAAAA 1599

Qy   135 AsnLysGlyLys 138
           :::
Db   1588 GACATGGGAAAA 1599

RESULT 12
US-08-747-221B-34
; Sequence 34, Application US/08747221B
; Patent No. 6063610
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; APPLICANT: Wisnewski, Nancy
; TITLE OF INVENTION: No. 606361O el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,221B
; FILING DATE: No. 6063610ember 12, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: PC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785

US-08-747-221B-34

Alignment Scores:
Pred. No.:
Score: 0.209
Percent Similarity: 80.50
Best Local Similarity: 38.41%
Query Match: 23.17%
DB: 11.26%
                                  3

Length: 1785
Matches: 38
Conservative: 25
Mismatches: 62
Indels: 39
Gaps: 8

US-09-701-618A-4 (1-141) x US-08-747-221B-34 (1-1785)

Qy   4 GluGluMetAsnGluAspTyrProValGlulleHisGluSerLeuThrAlaLeuGluSer 23
           :::||| |||::: |||:::
Db   1213 AAAGACAAAACGATGACATC-----------------------GAAGCCTATGTCCAACTAAAAGGC 1257

Qy   24 SerLeuGlyAlaValAspAspMetLeuThrMetMetAlaValSerArgAsnGluLeu 43
           ||| :::|||:::|||:::   :::||||||:::
Db   1258 GATGCTGGTTATATCTCCAAGGAATTACCGTACGTTACCTTGAAAGCCATATATTTTCAATGAAATC 1317

Db   1318 AAAAGAAATCCAACTGTTATTTGTTATAGGTTATTCAGATGATAACGTATTAGTGTTATAAAA 1377
           |||::: |||

Db   1378 AGTTATAATCTTGCCCTATCGATGGGGGGTTCCTTGCCAGGAGTTAGTCATGGTGATGATTA 1437

Qy   60 AlaTyrThrLeu-------------------------------------- 63

Db   1438 GGATATCTTTTGCAAACTCTTTGGATGTTCCTATTTTGGGGAAACAACGCACATTTCTATA-- 1497

Qy   64 --------AsnSerMetPheTrpValTyrlleuAlaThrGlnGlyValAsn-- 77
           ||| :::   :::
Db   1498 CCGCAAGAT------GCTATGCAGACTTCTGGAAAAGGATGTGTCAGGATCTGGAACCAAATTT 1551

Qy   78 ProLysGluHisProValLysGlnGluLeuGlyArg---IleArgValTyrMetAsnArg 96
           |||:::   ::: |||:::||||||:::
Db   1552 GTAAAGAATGGAAAAAACCTACATCAAACACTGAAGATGCATCATGTGAT--------ACA 1602

Qy   97 ValLys-----------GlulleThrAspLysLysAlaLeuThrProLysSerThrProLysValAla 114
           :: ||| :::
Db   1603 AAAAGACATTTAAACGACATTTTTGGGAACCATACAACGACGAGAACCAAAATATTTG 1662

Qy   115 SerArgPheValLysLysAlaLeuThrpGluProLysSerThrProLysValAla 134
           ||||||| |||||||
Db   1663 GACATGGGAAAA 1674

Qy   135 AsnLysGlyLys 138
           :::|||
Db   1663 GACATGGGAAAA 1674

RESULT 13
US-08-747-221B-35/c
; Sequence 35, Application US/08747221B
; Patent No. 6063610
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; APPLICANT: Wisnewski, Nancy
; TITLE OF INVENTION: No. 606361O el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,221B
; FILING DATE: No. 6063610ember 12, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: PC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

US-08-747-221B-35

Alignment Scores:

Pred. No.:

Score:	0.209
Percent Similarity:	80.50
Best Local Similarity:	38.41%
Query Match:	23.17%
	11.26%

Length:	1785
Matches:	38
Conservative:	25
Mismatches:	62
Indels:	39
Gaps:	8

US-09-701-618A-4 (1-141) x US-08-747-221B-35 (1-1785)

```
Qy      4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
          :::||||  |||  :::|||  |||  :|  |||  ::::
Db    573 AAAGAACAAAACGATGACATC------GAAGCCTATGTCCAACTAAAAGGC 529

Qy     24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValSerArgAsnGluLeu 43
          |||  ||||  ||  |  :||
Db    528 GATGCTGGTTATCTCCAAGGAAATTACCGTACCTTGAAAGCCATATTTTTCAATGAAATC 469

Qy     44 LeuGlnLys------------LeuAspProLeuGluGlnAlaLysValAspLeuValSer 59
                          :::|||  |||  |||  |||  :::
Db    468 AAAAGAAAATTCCAACTTGTATTTGTCAGATGATACGTATAGTGTATATATAAA 409

Qy     60 AlaTyrThrLeu------------------------------------------ 63
          |||  |||
Db    408 AGTTATATCTTGCCCTATCGATGGGGGTTCCTTGCCAGGAGTTAGTCATGGTGATGATTA 349

Qy     64 ------------AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
          ||||:|:::
Db    348 GGATATCTTTTTGCAAACTCTTGGATGTTCCTATTTTTGGGACAACGCACATTTCTATA 289

Qy     78 ProLysGluHisProValLysGlnIleLeuGluArg---IleArgValTyrMetAsnArg 96
          |||  |||  |||:|||  |||  :||:||||  ||||
Db    288 CCGCCAAGAT------------GCTATGCGAGACTCTGGAAAAGATGGTCAGGATCTCGGACCAATTTT 235

Qy     97 ValLys--------GluIleThrAspLysLysLysAlaLysLeuAspArgGlyAlaAla 114
          |||  |:||::  ::||  :||  |||  :|||
Db    234 GTAAAGAATGAGAAAACCTACATCAAACACTGAAGATGCATCATGTGAT-------ACA 184

Qy    115 SerArgPheValLysLysAlaLeuTrpGluProLysArgLysSerThrProLysValAla 134
          |||:|||  ||  |||
Db    183 AAAAGACATTTAAACGACATTTTTTGGGAACCATACAACGACGAAGAACCAAAAATATTTG 124

Qy    135 AsnLysGlyLys 138
          :::  ||||||
Db    123 GACATGGGAAAA 112
```

US-09-005-051-34

Alignment Scores:

Pred. No.:

Score:	0.209
Percent Similarity:	80.50
Best Local Similarity:	38.41%
Query Match:	23.17%
	11.26%

Length:	1785
Matches:	38
Conservative:	25
Mismatches:	62
Indels:	39
Gaps:	8

US-09-701-618A-4 (1-141) x US-09-005-051-34 (1-1785)

```
Qy      4 GluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
          :::||||  |||  :::|||  |||  :|  |||  ::::
Db   1213 AAAGAACAAAACGATGACATC------GAAGCCTATGTCCAACTAAAAGGC 1257

Qy     24 SerLeuGlyAlaValAspAspMetLeuLysThrMetMetAlaValSerArgAsnGluLeu 43
          |||  ||||  ||  |  :||
Db   1258 GATGCTGGTTATCTCCAAGGAAATTACCGTACCTTGAAAGCCATATTTTTCAATGAAATC 1317

Qy     44 LeuGlnLys------------LeuAspProLeuGluGlnAlaLysValAspLeuValSer 59
                          :::|||  |||  |||  |||  :::
Db   1318 AAAAGAAAATTCCAACTTGTATTTGTCAGATGATACGTATAGTGTATATATAAA 1377

Qy     60 AlaTyrThrLeu------------------------------------------ 63
          |||  |||
Db   1378 AGTTATATCTTGCCCTATCGATGGGGGTTCCTTGCCAGGAGTTAGTCATGGTGATGATTA 1437

Qy     64 ------------AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
          ||||:|:::
Db   1438 GGATATCTTTTTGCAAACTCTTGGATGTTCCTATTTTTGGGACAACGCACATTTCTATA 1497

Qy     78 ProLysGluHisProValLysGlnIleLeuGluArg---IleArgValTyrMetAsnArg 96
          |||  |||  |||:|||  |||  :||:||||  ||||
Db   1498 CCGCCAAGAT------------GCTATGCGAGACTCTGGAAAAGATGGTCAGGATCTCGGACCAATTTT 1551

Qy     97 ValLys--------GluIleThrAspLysLysLysAlaLysLeuAspArgGlyAlaAla 114
          |||  |:||::  ::||  :||  |||  :|||
Db   1552 GTAAAGAATGAGAAAACCTACATCAAACACTGAAGATGCATCATGTGAT-------ACA 1602

Qy    115 SerArgPheValLysLysAlaLeuTrpGluProLysArgLysSerThrProLysValAla 134
          |||:|||  ||  |||
Db   1603 AAAAGACATTTAAACGACATTTTTTGGGAACCATACAACGACGAAGAACCAAAAATATTTG 1662

Qy    135 AsnLysGlyLys 138
          :::  ||||||
Db   1663 GACATGGGAAAA 1674
```

```
;  APPLICANT: Silver, Gary W.
;  APPLICANT: Wisnewski, Nancy
;  TITLE OF INVENTION: No. 6291222e1 Carboxylesterase Nucleic Acid
;  TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
;  NUMBER OF SEQUENCES: 66
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Carol Talkington Verser, Ph.D.
;  ADDRESSEE: Heska Corporation
;  STREET: 1825 Sharp Point Drive
;  CITY: Fort Collins
;  STATE: Colorado
;  COUNTRY: USA
;  ZIP: 80525
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: Windows 95
;  SOFTWARE: WordPerfect for Windows, Version 7.0
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/09/005,051
;  FILING DATE:
;  CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: 08/747,221
;  FILING DATE: No. 6291222ember 12, 1996
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Verser, Carol Talkington
;  REGISTRATION NUMBER: 37,459
;  REFERENCE/DOCKET NUMBER: FC-1
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: 970/493-7272
;  TELEFAX: 970/484-9505
;  INFORMATION FOR SEQ ID NO: 35:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 1785 nucleotides
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: cDNA
US-09-005-051-35

Alignment Scores:
Pred. No.:
Score:                    0.209
Percent Similarity:       80.50
Best Local Similarity:    38.41%
Query Match:              23.17%
DB:                       11.26%
                          3

US-09-701-618A-4 (1-141) x US-09-005-051-35 (1-1785)

                                              Length:        1785
                                              Matches:         38
                                              Conservative:    25
                                              Mismatches:      62
                                              Indels:          39
                                              Gaps:             8
```

```
Qy     4 GluGlumetAsnGluAspTyrProValGluIleHisGluSerLeuThrAlaLeuGluSer 23
           :::||||||||:::||||||||||||        :::||||||||||
Db   573 AAAGACAAAACGATGACATC----------------GAAGCCTANGTCCAACTAAAAGGC 529

Qy    24 SerLeuGlyAlaValAspAspMetLeuLysThrMetAlaValSerArgAsnGluLeu 43
           ||||            ::::    ||:::||||||::||||:||
Db   528 GATGCTGGTTATCTCCAAGGAATTTACCGTACCTTGAAAGCCATATTTTTCAATGAAATC 469

Qy    44 LeuGlnLys----------LeuAspProLeuGluGlnAlaLysValAspLeuValSer 59
           :::                ||||
Db   468 AAAAGAAAAATTCCAACTTGTATTTGTATAGGTTATCAGATGATAGATACCGTATAGTGTTATATAAA 409

Qy    60 AlaTyrThrLeu---------- 63
           :||||:::||||
Db   408 AGTTATATCTTGCCTATCGATGGGGTTCCTTGCCAGGAGTTAGTCATGTGATCGAATTTA 349

Qy    64 ------------AsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsn--- 77
             ||||||:||:
Db   348 GGATATCTTTTGCAAACTCTTGGATGTTCCTATTTGGGAACAACGCACATTCTATA 289

Qy    78 ProLysGluHisProValLysGlnGluLeuGluArg---IleArgValTyrMetAsnArg 96
           ||| ::||||:::::       :::||||||:::::       |||
Db   234 GTAAAGAATGGAAAACCTACATCAATCAAACACTGAAGAATGCATCATGTGAT--------ACA 184
```

```
Db   288 CCGCAAGAT--------GCTATCGAGACTCTGAAAAGAATGGTCAGAGATCTGACCAATTT 235
           ||||||||
Qy    97 ValLys------GluIleThrAspLysGlyLysSerAlaAlaLysLeuAspArgGlyAlaAla 114
           :::           |||           |||
Db   234 GTAAAGAATGGAAAACCTACATCAATCAAACACTGAAGAATGCATCATGTGAT--------ACA 184
Qy   115 SerArgPheValLysLysAlaLeuTrpGluProLysGlyLysSerThrProLysValAla 134
           ||||||            |||||||
Db   183 AAAAGACATTTAAACGACCATTTTTTGGGACCATACAACGACGAAGAACCAAAAATATTTG 124
Qy   135 AsnLysGlyLys 138
           :::  ||||||
Db   123 GACATGGGAAAA 112
```

```
Search completed: January  5, 2005, 16:12:48
Job time : 76 secs
```

This Page Blank (uspto)

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: January 5, 2005, 13:19:23 ; Search time 339.5 Seconds
                       (without alignments)
                       2345.815 Million cell updates/sec

Title: US-09-701-618A-4
Perfect score: 715
Sequence: 1 MAGEMNEDYPVEIHESLTA..........LWEPRKSTPKVANKGKSKH 141

Searched: 4176236 seqs, 282412795 residues

Scoring table: BLOSUM62

	Xgapop	10.0	;	Xgapext	0.5
	Ygapop	10.0	;	Ygapext	0.5
	Fgapop	6.0	;	Fgapext	7.0
	Delop	6.0	,	Delext	7.0

Total number of hits satisfying chosen parameters: 8352472

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09701618/runat_05012005_085840_24491/app_query.fasta_1.654
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09701618 @CGN_1_1_879 @runat_05012005_085840_24491
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Published_Applications_NA.*
  1: /cgn2_6/ptcdata/1/pubpna/US07_PUBCOMB.seq:*
  2: /cgn2_6/ptcdata/1/pubpna/PCT_NEW_PUB.seq:*
  3: /cgn2_6/ptcdata/1/pubpna/US09C_PUBCOMB.seq:*
  4: /cgn2_6/ptcdata/1/pubpna/US09B_PUBCOMB.seq:*
  5: /cgn2_6/ptcdata/1/pubpna/US09A_PUBCOMB.seq:*
  6: /cgn2_6/ptcdata/1/pubpna/US09_NEW_PUB.seq:*
  7: /cgn2_6/ptcdata/1/pubpna/US08_PUBCOMB.seq:*
  8: /cgn2_6/ptcdata/1/pubpna/PCTUS_PUBCOMB.seq:*
  9: /cgn2_6/ptcdata/1/pubpna/US07_NEW_PUB.seq:*
 10: /cgn2_6/ptcdata/1/pubpna/US06_PUBCOMB.seq:*
 11: /cgn2_6/ptcdata/1/pubpna/US06_NEW_PUB.seq:*
 12: /cgn2_6/ptcdata/1/pubpna/US09_NEW_PUB.seq:*
 13: /cgn2_6/ptcdata/1/pubpna/US10B_PUBCOMB.seq:*
 14: /cgn2_6/ptcdata/1/pubpna/US10A_PUBCOMB.seq:*
 15: /cgn2_6/ptcdata/1/pubpna/US10C_PUBCOMB.seq:*
 16: /cgn2_6/ptcdata/1/pubpna/US10D_PUBCOMB.seq:*
 17: /cgn2_6/ptcdata/1/pubpna/US10E_PUBCOMB.seq:*
 18: /cgn2_6/ptcdata/1/pubpna/US10_NEW_PUB.seq:*
 19: /cgn2_6/ptcdata/1/pubpna/US11_NEW_PUB.seq:*
 20: /cgn2_6/ptcdata/1/pubpna/US60_NEW_PUB.seq:*
 21: /cgn2_6/ptcdata/1/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match Length	DB	ID	Description
1	704	98.5	426	15	US-10-414-692-42	Sequence 42, Appl
2	646	90.3	1172	10	US-09-873-367C-318	Sequence 318, App
3	634	88.7	586	15	US-10-029-386-1780	Sequence 1780, Ap
c 4	538.5	75.3	457	16	US-10-242-535A-49006	Sequence 49006, A
c 5	538.5	75.3	457	16	US-10-085-783A-49006	Sequence 49006, A
c 6	235	32.9	204	15	US-10-029-386-15480	Sequence 15480, A
7	141	19.7	1302	16	US-10-424-599-94086	Sequence 137946, A
8	133	18.6	1099	18	US-09-923-876-2854	Sequence 137946, A
c 9	125.5	17.6	1302	16	US-10-424-599-94086	Sequence 94086, A
10	119.5	16.7	874	16	US-09-908-975-7045	Sequence 1617, Ap
11	119.5	16.7	1048	17	US-10-767-701-10044	Sequence 10044, A
12	116	16.2	630	17	US-10-437-963-24791	Sequence 24791, A
13	103	14.4	476	13	US-10-027-632-299522	Sequence 299522,
c 14	103	14.4	476	13	US-10-027-632-299522	Sequence 299522,
15	100.5	14.1	2823	17	US-10-437-963-45565	Sequence 45565, A
16	98	13.7	65	10	US-09-908-975-262240	Sequence 262240, A
17	94	13.1	227	9	US-09-923-876-2854	Sequence 2854, Ap
18	94	13.1	227	10	US-09-923-876-2854	Sequence 2854, Ap
19	91	12.7	688	16	US-09-923-876-299034	Sequence 99034, A
20	89	12.4	60	10	US-09-908-975-7045	Sequence 7045, Ap
21	89	12.4	527	17	US-10-767-701-21423	Sequence 21423, A
22	88.5	12.4	53	10	US-10-767-701-53	Sequence 53, Appl
23	87.5	12.2	1685	17	US-10-115-635-53	Sequence 4984, Ap
24	86.5	12.1	859	17	US-10-437-963-4984	Sequence 4984, Ap
25	86.5	12.1	1527	16	US-10-282-122A-27913	Sequence 27913, A
26	84.5	11.8	3750	9	US-09-938-842A-1662	Sequence 1662, Ap
27	84	11.8	1035	11	US-09-938-842A-1662	Sequence 1662, Ap
28	84	11.7	1035	10	US-09-906-179A-142	Sequence 142, App
29	84	11.7	1035	10	US-09-906-179A-143	Sequence 143, App
30	84	11.7	1960	17	US-10-437-963-65096	Sequence 65096, A
31	83.5	11.7	13734	16	US-10-287-226-303	Sequence 303, App
32	83	11.6	3222	10	US-09-882-876-2595	Sequence 595, App
33	83	11.6	236	9	US-09-923-876-2728	Sequence 2728, Ap
34	83	11.6	2074	18	US-09-923-876-2728	Sequence 2728, Ap
35	82.5	11.5	7100	19	US-10-425-115-3612	Sequence 3612, Ap
36	82	11.5	414	9	US-09-932-183A-1	Sequence 1, Appli
37	81.5	11.4	1350	16	US-10-960-352-6162	Sequence 6162, Ap
38	81.5	11.4	2080	18	US-10-282-122A-22828	Sequence 22828, A
39	81.5	11.4	1710	16	US-10-425-115-47599	Sequence 47599, A
40	80.5	11.3	1785	16	US-10-678-521-33	Sequence 33, Appl
c 41	80.5	11.3	1785	16	US-10-678-521-34	Sequence 34, Appl
42	80.5	11.3	1785	16	US-10-678-521-35	Sequence 35, Appl
43	80.5	11.3	2457	17	US-10-437-963-64138	Sequence 64138, A
c 44	80.5	11.3	2801	16	US-10-678-521-30	Sequence 30, Appl
45	80.5	11.3	3285	15	US-10-032-585-6554	Sequence 6554, Ap

ALIGNMENTS

RESULT 1
US-10-414-692-42

; Sequence 42, Application US/10414692
; Publication No. US20030228607A1
; GENERAL INFORMATION:
; APPLICANT: X-Ceptor Therapeutics, Inc.
; TITLE OF INVENTION: Screening method and modulators having an improved therapeutic
; TITLE OF INVENTION: profile
; FILE REFERENCE: 8012-002-US
; CURRENT APPLICATION NUMBER: US/10/414,692
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 60/372,650
; PRIOR FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-414-692-42

RESULT 2

US-09-873-367C-318

Sequence 318, Application US20030165839A1

Publication No. US20030165839A1

US-09-873-367C-318

Alignment Scores:

Pred. No.:		
Score:	1.78e-79	Length: 426
	704.00	Matches: 139
Percent Similarity:	98.58%	Conservative: 0
Best Local Similarity:	98.58%	Mismatches: 2
Query Match:	98.46%	Indels: 0
DB:	15	Gaps: 0

US-09-701-618A-4 (1-141) x US-10-414-692-42 (1-426)

```
Qy    1 MetAlaGlyGlyGluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAla 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGCAGGTGAAGAAATGAATGAAGATTATCCCGTAGAAATTCACGAGTCTTTAACAGCC 60

Qy   21 LeuGluSerSerLeuGlyAlaValAspAspMetLeuLysThrMetAlaValSerArg 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTGGAGAGCTCCCTGGGTGCTGTGGATGACATGCTGAAGACCATGGCTGTGTTCTAGA 120

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnLnAlaLysValAspLeuValSerAla 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AATGAGTTGTTGCAGAAGTTGGACCCATTGCAAACAGCCAAAGTGGATTTAGTTTCTGCA 180

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TACACCTTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAGGAGTTAATCCCAAAGAG 240

Qy   81 HisProValLysGlnLeuGluLeuGluArgIleArgValThrMetAsnArgValLysGlyIle 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CATCCAGTGAAGCAGCAGAACTCGACAGTCTACATGAACAGAGTTAAAGAAATA 300

Qy  101 ThrAspLeuGlyLysGlyLysAlaAlaLysLysLeuAspArgGlyAlaAlaSerArgPheValLysLys 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ACAGACAGAAGGAGAAAGGCTGCCAAGCTGGACAGAGGTGCTGCTTCAGAATTTGTCAAGAAG 360

Qy  121 AlaLeuTrpGluProGlyArgGlySerThrProLysValAlaLysAsnLysGlyLysSerLys 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GCAACTCTGGGAACCCAAAGGACACACCAAAAGTGGCTAAATAAAGGGGAAAAACCAA 420

Qy  141 His 141
        |||
Db  421 CAC 423
```

US-09-701-618A-4 (1-141) x US-09-873-367C-318 (1-1172)

Alignment Scores:

Pred. No.:		
Score:	1.55e-71	Length: 1172
	646.00	Matches: 126
Percent Similarity:	95.00%	Conservative: 7
Best Local Similarity:	90.00%	Mismatches: 7
Query Match:	90.35%	Indels: 0
DB:	10	Gaps: 0

US-09-701-618A-4 (1-141) x US-09-873-367C-318 (1-1172)

```
Qy    1 MetAlaGlyGlyGluGluMetAsnGluAspTyrProValGluIleHisGluSerLeuThrAla 20
        |||||||||||||||||||||||||||||||||| ::||||||||||||||||||||||
Db  118 ATGGCAGGTGAAGAAATGAATGAAGATTATCCAGTAGAAATTCACGAGTATTGTCAGCG 177

Qy   21 LeuGluSerSerLeuGlyAlaValAspAspMetLeuLysThrMetAlaValSerArg 40
        ||||||||||||||||||||||||| ::|||||||||||||||||||||||||||
Db  178 TTTGAGAATTCCATTGGTGTGCTGTGGATGACATGCTGAAGACCATGGCTGTGTTCTAGA 237

Qy   41 AsnGluLeuLeuGlnLysLeuAspProLeuGlnLnAlaLysValAspLeuValSerAla 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  238 AATGAGTTGTTGCAGAAGTTGGACCCATTGCAAACAGCCAAAGTGGATTTAGTTTCTGCA 297

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGlu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  298 TACACCTTAAATTCAATGTTTTGGGTTTATTTGGCAACTCAAGGAGTTAATCCCAAAGGA 357

Qy   81 HisProValLysGlnLeuGluLeuGluArgIleArgValThrTyrMetAsnArgValLysGlyIle 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  358 CATCCAGTAAAACAGGAGTTGGAAAGAATCAGAGTTATATAAGAACAGAGTCAAGGAAATA 417

Qy  101 ThrAspLeuGlyLysGlyLysAlaAlaLysLysLeuAspArgGlyAlaAlaSerArgPheValLysLys 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  418 ACAGACAGAAGGAGAAAGGCTGCCAAGCTGGACAGAGGTGCTGCTTCAGAATTTGTAAAAAT 477

Qy  121 AlaLeuTrpGluProGlyArgGlySerThrProLysValAlaLysAsnLysGlyLysSerLys 140
        |||||||||||||||||||||||||||| :::||||||||||||||||||||||
Db  478 GCCTTCTGGGAACCAAAAATCGAAAAAATCCATCAAAAGTTGCCAATAAAGGGAAAAGTAAA 537
```

RESULT 3

US-10-029-386-1780/c

Sequence 1780, Application US/10029386

Publication No. US20030194704A1

US-10-029-386-1780

Alignment Scores:

Pred. No.:	1.95e-70	Length: 586

Score: 634.00   Matches: 124
Percent Similarity: 94.24%   Conservative: 7
Best Local Similarity: 89.21%   Mismatches: 8
Query Match: 88.67%   Indels: 0
DB: 15   Gaps: 0

US-09-701-618A-4 (1-141) x US-10-029-386-1780 (1-586)

Qy   2 AlaGlyGlyGluGluMetAsnGluAspThrTyrProValGlyIleHisGluSerLeuThrAlaLeu 21
        ||| :|||||||||||||||||||||||||||||||||||||:||||
Db 515 GCAGGTGAAGAAATAATGAAGACTATCCAGTAGAAATTCACGGCGTTT 456

Qy 22 GluSerSerLeuGlyAlaValAlaAspAlaspMetLeuLysThrMetAlaValSerArgAsn 41
        ||||||||||||||||||||| :||||||||||||||||||:||
Db 455 GAGAATTCCATTGGTCGTCGTGGATGCTGAGGACCACGATGTCTGTTTTCTAGAAAT 396

Qy 42 GluLeuLeuGlnLysGluAsnProLeuGlnIleAlaLysValAspLeuValSerAlaTyr 61
        ||||||||||||||||||||||||||||||||||||||||||||
Db 395 GAGTTGTTGCAGAAGTTGGATCCACTTGAACAAGCAAAAGTGGATTTGGTTTCTGCATAC 336

Qy 62 ThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnAspValValAsnProLysGluHis 81
        ||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db 335 ACATTAAATTCAATGTTTTGGGTTTATTTGGCAACCCAAGGATGTTGTTAATCCTAAGGAACAT 276

Qy 82 ProValLysGlyGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGlyIleThr 101
        ||||||||||||||||||||||||||||||||||||||||||||||||:||
Db 275 CCAGTAAAACAGGAATTCAGAGTAATCAGAGAGTATATGAACAGAGTTAATCCTAAGGAACA 216

Qy 102 AspLysLysLysLysAlaAlaLysLeuAspArgGlyAlaAlaSerArgPheValLysLysAla 121
         |||||||||||||||||||||||||||||||||||||||||||:|||
Db 215 GACAAGAAAAAGGCTGCCAAGCTGGACAGAGGTGCAGCTTCAGAGTTTGTAAAAAATGCC 156

Qy 122 LeuTrpGluProLysArgLysSerThrProLysValAlaAlaAsnLysGlyLysSerLys 140
         |||||||||||||:|||||:::||
Db 155 CTCTGGGAACCGAAAAGATCCATAAAAGTTGCCGCAATAAAGGGAAAAGGTAA 99

FEATURE:
NAME/KEY: misc_feature
LOCATION: (381)..(381)
OTHER INFORMATION: n is a, c, g, or t

US-09-701-618A-4 (1-141) x US-10-242-535A-49006 (1-457)
DB:

Alignment Scores:
Pred. No.:   1.63e-58   Length: 457
Score: 538.50   Matches: 117
Percent Similarity: 87.14%   Conservative: 5
Best Local Similarity: 83.57%   Mismatches: 18
Query Match: 75.31%   Indels: 3
DB: 16   Gaps: 0

Qy   1 MetAlaGlyGlyGluGluMetAsnGluAspThrTyrProValGlyIleHisGluSerLeuThrAla 20
         ||||||||||||||||||||||||||||||||||||||||||||||:|||
Db   30 ATGGCAGGTGAAGAAATGAATGAAGACTATCCAGTAGAAATTCACGGTAGTTTGNCAGCG 89

Qy   21 LeuGluSerSerLeuGlyAlaValAlaAspAlaspMetLeuLysThrMetAlaValSerArg 40
         ||||||||||||||:||||||||||||||||||||||||||||||
Db   90 TTTGAGAATTCCATTGGGTCGTCTGTGGATGCTGAGGACCATGATGCCGTGTTTCTAGA 149

Qy   41 AsnGluLeuLeuGlnLysGluAsnProLeuGlnIleAlaLysValAspLeuValSerAla 60
         ||||||||||||||||||||||:||::||||||||||||||||||
Db 150 AATGAGTTGTTGCAGAAGTTGGATCCAACAGCAAAAGTGGATTTGGTTTCTGCA 209

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValValAsnProLysGlu 80
         |||||||||||||||||||||||:||||||||||:|||||||||||||
Db 210 TACACATTAAATTCAATGTTTTGGGNTTATTTGGCAACCCAAGGAGTNAATCCTAAGGAA 269

Qy   81 HisProValLysGlyGlnGluLeuGluArgIleArgValTyrMetAsnArgValLysGluIle 100
         |||||||||||||||||||||||||||||||||||:||
Db 270 CATCCAGTAAAACAGGAATTCAGAGAGTATATGAACAGAGTAAACAGAGGGAAATA 329

Qy 101 ThrAspLysLysLysLysAlaAlaLysLeuAspArgGlyAlaAlaAlaSerArgPheValLysLys 120
         ||||||||||||| ||||||||||||||||||||||||||||:|||
Db 330 ACAGACAAGAAAAAGGCT--GGCAAGCTGGACAGAGGTGCAGCTTCAGAGTTTGTNAAAAAAT 388

Qy 121 AlaLeuTrpGluProLysArgLysSerThrProLysValAlaAlaAsnLysGlyLysSerLys 140
         ||| ||||||||||||||:||||:||:::|| |||
Db 389 GCCCT-TGGGAACCGAAACCGAAAAAATGCATCAAAAAGT--GCCAATAAAGGGAAAAGGAAA 445

; LOCATION: (84)..(84)
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (235)..(235)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (381)..(381)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-49006
DB:

Alignment Scores:
Pred. No.:
Score: 1.63e-58
Percent Similarity: 87.14%   Length: 457
Best Local Similarity: 83.57%   Matches: 117
Query Match: 75.31%   Conservative: 5
                                        Mismatches: 18
                                        Indels: 3
                                        Gaps: 0

US-09-701-618A-4 (1-141) x US-10-085-783A-49006 (1-457)

Qy   1 MetAlaGlyGluGluMetAsnGluAspThrProValGluIleHisGluSerLeuThrAla 20
        ||||||||||||||||||||||||||||||||||||::||:||
Db   30 ATGGCAGGTGAAGAAATTAATGNAGACTATCCAGTAGAAATTCACCAGTATTGNCAGCG 89

Qy   21 LeuGluSerSerLeuGlyAlaValAlaAspAspMetLeuLysThrMetAlaValSerArg 40
        ||::|||||||
Db   90 TTTGAGAANTTCCATTTGGTGCTGTGGAGAGATGGCTGAAGACCATGATGCCTGTTTCTTAGA 149

Qy   41 AsnGluLeuLeuGlnIleLysLeuAspProLeuGlnIleAlaLysValAlaValSerAla 60
Db 150 AATGAGTTGTTGCAGAAGAGTTGGATCCACTTGAACAGACAAAAAGTGGGATTTGGTTTCTGCA 209

Qy   61 TyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyValAlaAsnProLysGlu 80
Db 210 TACACATTAAATTCAAATGTTTGGGNTTATTTGGCAACCCAAGGAGTTAATCCTAAGGA 269

Qy   81 HisProValLysGlyGlnLeuArgIleArgValTyrMetAsnArgValLysGlyGluIle 100
Db 270 CATCCACAGTAAAACAGGAATTCGAGAAATCAGAGTATATGAACAGAGCAAGGGAATA 329

Qy 101 ThrAspLysGlyLysGlyAlaLeuAlaLysLeuAspArgGlyAlaAlaSerArgPheValLys 120
Db 330 ACAGACAAGAAGAAAAGCT-GGCAAGCTGGGACAGAGAGGTGCCAGCTTCAAGATTTGNAAAAAAT 388

Qy 121 AlaLeuTrpGluProLysArgLysSerThrProLysGlnValAlaAlaAsnLysGlyLysSerLys 140
Db 389 GCCCT-TGGGACCAAACCGAAAAATGCATCAAAAAGT--GCCAATAAAGGAAAAAGGAAA 445

RESULT 6
US-10-029-386-15480/c
; Sequence 15480, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;   APPLICANT: Penn, Sharron G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;   TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
;   FILE REFERENCE: ABOMICA-X-2
;   CURRENT APPLICATION NUMBER: US/10/029,386
;   CURRENT FILING DATE: 2001-12-20
;   NUMBER OF SEQ ID NOS: 34288
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;   SEQ ID NO 15480
;   LENGTH: 204
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO CHR10.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.1
; OTHER INFORMATION: NT HIT: gi14743475, EVALUE 1.00e-111
; OTHER INFORMATION: EST HUMAN HIT: AV763046.1, EVALUE 1.00e-108
; OTHER INFORMATION: SWISSPROT HIT: Q9NQ30, EVALUE 4.20e+00
US-10-029-386-15480
DB:

Alignment Scores:
Pred. No.:
Score: 1.24e-20
Percent Similarity: 90.57%   Length: 204
Best Local Similarity: 88.68%   Matches: 47
Query Match: 32.87%   Conservative: 1
                                        Mismatches: 5
                                        Indels: 0
                                        Gaps: 0

US-09-701-618A-4 (1-141) x US-10-029-386-15480 (1-204)

Qy   88 GluArgIleArgValTyrMetAsnArgValLysGlyGluIleThrAspLysGlyLysAlaAla 107
        ||||||||||||||||||||||||||||||||||||||||
Db 203 GAAAGAATCAGAGTATATGAACAGAGTCAAGGGAGAAATAACAGACAAGAAAAAGGCTGGC 144

Qy 108 LysLeuAspArgGlyAlaAlaSerArgPheValLysAlaLeuTrpGluProLysArg 127
        |||||||||||||||||||||||||||||||||||
Db 143 AAGCTGGACAGAGAGGTGCCAGCTTCAAGATTTGTAAAAAATGCCCTGGGACCAAACCG 84

Qy 128 LysSerThrProLysGlnValAlaAlaAsnLysGlyLysSerLys 140
        :::|
Db   83 AAAAATGCATCAAAAAGTTGCCAATAAAGGAAAAAGGGTAAA 45

RESULT 7
US-10-424-599-94086
; Sequence 94086, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa Thomas J
;   APPLICANT: Kovalic David K
;   APPLICANT: Zhou Yihua
;   APPLICANT: Cao Yongwei
;   TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53223)B
;   CURRENT APPLICATION NUMBER: US/10/424,599
;   CURRENT FILING DATE: 2003-04-28
;   NUMBER OF SEQ ID NOS: 285684
;   SEQ ID NO 94086
;   LENGTH: 1302
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_55972C.1
US-10-424-599-94086

Alignment Scores:
Pred. No.:
Score: 1.24e-07
Percent Similarity: 52.90%   Length: 1302
Best Local Similarity: 24.64%   Matches: 34
Query Match: 19.72%   Conservative: 39
                                        Mismatches: 57
                                        Indels: 8
                                        Gaps: 2

US-09-701-618A-4 (1-141) x US-10-424-599-94086 (1-1302)

Qy   11 ProValGluIleHisGluSerLeuThrAlaLeuGluSerSerLeuGlyAlaValAlaAspAsp 30
        ||| :::||||||| :::
Db 314 CCGGAACCGGTTATGGAATTCAGTCAACAACCACTCTCTCCAACCTTCAACAACTCGGAACG 373

Qy   31 MetLeuLysThrMetAlaValSerArgAsnGluLeuLeuGlnIleLysLeuAspProLeu 50
        :::||||| ||| ::: :::
Db 374 CACTTAACGAGTTTTTGTCTCTCTCCGACCCTCAAACCCTCTCCAAATGCCCCTCTT 433

Qy   51 GluGlnIleAlaLysValAlaAspLeuValSerAlaTyrThrLeuAsnSerMetPheTrpValTyr 70

```
Db  434 CAACCGGCGTCATTCCTCCTTCGCCCAAAGTCACATCTACCCTTCGTTGCATTGAAC 493
            ::: :::                                        :::
Qy   71 LeuAlaThrGlnGlyValAsnProLysGlnGluHisProValLysGlnGluArgIle 90
            ||| ||| |||
Db  494 CTAAGGTGTACCGGCGGGGTCCACCCAGATGATCATCCCATCAAATCAGAGCTTGATAGGGTA 553

Qy   91 ArgValTyrMetAsnArgValLysGluIleThrAspLysAspLysAla--------- 106
            :::::|||
Db  554 AGCTTGTACGAGGCAAAACTGGAACGTTTGCTTCTGGATTGAGTAAAGCTTCCATTGGGACCT 613

Qy  107 ---AlaLysLeuAspArgGlyAlaAlaAlaSerArgPheValLysLysAlaLeuTrpGluPro 125
            ||| || |::::
Db  614 TCTACTACCTTGAATTATCAGGCAGCCAGCTTACACGCTTCCATTAGAACACTCTCGTCCTGACCT 673

Qy  126 LysArgLysSerThrProLysValAlaAsn----------LysGlyLysSerLys 140
            |||||:::
Db  674 AACAACAGAGACGAGGGAAAATATCAGGAACACATTAGTAGGAGGGGAGGAGACCAAA 727
```

RESULT 8
US-10-425-115-137946
; Sequence 137946, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; OTHER INFORMATION: Clone ID: MRT4577_57283C.1
; SEQ ID NO 137946
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NUMBER OF SEQ ID NOS: 369326

US-10-425-115-137946

Alignment Scores:
Pred. No.:

Score:	133.00	Length: 1099
	1e-06	Matches: 41
Percent Similarity:	45.03%	Conservative: 27
Best Local Similarity:	27.15%	Mismatches: 51
Query Match:	18.60%	Indels: 32
DB:	18	Gaps: 5

US-09-701-618a-4 (1-141) x US-10-425-115-137946 (1-1099)

```
Qy   18 LeuThrAlaLeuGluSerSerLeuGly----------AlaValAspAspMetLeuLysThr 34
            ||| ||
Db  170 GTCTCAGCTTGCCGAGGAGAACGCTCGGGATATACCGGAAATCCGTAGGGGGGACCATTCTCCCAT 229

Qy   35 MetMetAlaValSerArg----------AsnGluLeuLeuGlnLysLeuAspProLeuGluGln 52
            ::: |||
Db  230 CTGCTCCGCGGCGGCGGCGGGGGGGGGGGAGCCGTGGCCAGCTCCCACCCCTCTCCGG 289

Qy   53 AlaLysValAspLeuValSerAlaTyrThrLeuAsnSerMetPheTrpValTyrLeuAla 72
            ::: ||:
Db  290 GCGCGGCGCTTTCCTTGGCGCAGGCCGCGACCTCCCTTCCTCGGAGTTCGTTTAAGG 349

Qy   73 ThrGlnGlyValAlaAsnProLysGlnHisProValLysGlnGluLeuGluArgIleArgVal 92
            ||| ||
Db  350 TGTTCGGGAGTTGACCCTGCGACGGCGACACCCCATCAGAAGAGGAGTTTGAAAGGTTAAGCCTTA 409

Qy   93 TyrMetAsnArgValLysGluIleThrAspLysAspLysAla----------Ala 107
            :::::|||
Db  410 ATGCAGGAGAAGTAAATCAATTTGAGAACTGGGACAAAAGCACACTTCGCCCTTCTACT 469

Qy  108 LysLeuAspArgGlyAlaAlaAlaSerArgPheValLysLysAlaLeu---------- 122
```

```
Db  470 ACACTAAATACAAGAGCAGCAGCAAGGTTCATTGGACACTCACTTTCCCATCTGACATCC 529
            ||||||::: ::|||||||||
Qy  123 ------------------------------------------------------TrpGlu---Pro 125
Db  530 GATCAGAAGAGAGGAGCAGTGCATGCAAATAAGTAGGAGGAGAAAGGCGGAGTTGGTTGTCGGGCAG 589

Qy  126 LysArgLysSerThrProLysValAlaAsnLys 136
            |||||||
Db  590 AAGAGAAAAGCCTGAACCTTCAGTAGTAGAAAAGAAG 622
```

RESULT 9
US-10-424-599-94086/c
; Sequence 94086, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 94086
; LENGTH: 1302
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_55972C.1
US-10-424-599-94086

Alignment Scores:
Pred. No.:

Score:	125.50	Length: 1302
	1.13e-05	Matches: 26
Percent Similarity:	59.78%	Conservative: 29
Best Local Similarity:	28.26%	Mismatches: 32
Query Match:	17.55%	Indels: 5
DB:	16	Gaps: 1

US-09-701-618A-4 (1-141) x US-10-424-599-94086 (1-1302)

```
Qy   36 MetAlaValSerArgAsnGluLeuLeuGlnLysLeuAspProLeuGluGlnAlaLysVal 55
            ::::::||| ::|
Db  287 TTGTCTCTCCGACCCTCAAACCTTCTCCCAAATCGCCCCTCTTCCAAACGGGGCTCATTCC 228

Qy   56 AspLeuValSerAlaTyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGly 75
            :::
Db  227 CTCTCATCCTCGCCAAAGTCACATCTACCCTTCTGCAGTCATTGAACGTAAGGTGTACCGGG 168

Qy   76 ValAsnProLysGlnHisProValLysGlnGluLeuGluArgIleArgValTyrMetAsn 95
            |||::::|||||:::
Db  167 GTCCACCCAGATGATCATCCCATCAAATCAGAGCTTGATAAGCTTGTACGAGGAC 108

Qy   96 ArgValLysGluIleThrAspLysAspLysAla----------AlaLysLeuAsp 110
            |||:::
Db  107 AAACTGGAACGTTTGCTTCTTGATTTGAGTAAAGCTTCCATTGCGACCTTCTACTTGAAT 48

Qy  111 ArgGlyAlaAlaAlaSerArgPheValLysLysAlaLeu 122
            :::
Db   47 TATCAGGCAGCTACACGCTTCATCAGAACACTCTCTG 12
```

RESULT 10
US-10-425-114-1617
; Sequence 1617, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E

```
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1617
; LENGTH: 874
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700161056_FLI
US-10-425-114-1617

Alignment Scores:
Pred. No.:
Score:                        3.72e-05
Percent Similarity:           119.50
Best Local Similarity:        43.31%
Query Match:                  26.77%
DB:                           16.71%
                              16

US-09-701-618A-4 (1-141) x US-10-425-114-1617 (1-874)

Qy   37 AlaValSerArgAlaArgAsnGluLeuLeuGlnLysLeuLeuAspProLeuGlnIleAlaLysValAlaAsp 56
        |||     :::                       :|||     ||||
Db   31 GCGGCGGTCGCGGCGACCCCGAGCCCGTGGCCCAGCTTCCACCCTTTCTCCGGCGCGCGCTTTC 90

Qy   57 LeuValSerAlaThrTyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThrGlnGlyVal 76
        ||||           |||     |||||    |||     ||||
Db   91 CTTGTCGCCGTGGCGCTACGGGCCGACCTCCTTCTCGGAGGTTCGTTCGTTGTTTCGGGAGTT 150

Qy   77 AsnProLysGluHisProValLysGlnGluGluGlnLysGluArgIleArgValTyrMetAsnArg 96
        :::            :||:||                :::
Db  151 GACCCTGACGAGCACCCCATCAGAAAGGAGTTTGAAAGGTTAAGCCTAATCGACGAGGAAG 210

Qy   97 ValLysGluIleThrAspLysLysLysLysAla--------------------AlaLysLeuAspArg 111
        :::      |||                                            |||
Db  211 TTAAATCAATTTGAGAACTGGGACAAAGCACCACACTTCGCCCTTCTACTACACTAAATACA 270

Qy  112 GlyAlaAlaSerArgPheValLysLysAlaLeu------------------------ 122
        |||||     ||||||:::|||                             ||||
Db  271 CAAGCAGCAGCAAGGTTCATTGGACACTCACTTTCCATCTGACATCTGATCAGAAGAGG 330

Qy  123 --------------------------------------TrpGlu---ProLysArgLysSer 129
                                                |||        |||||
Db  331 AGCATGCATGAAATAAGTAGGAGGAAAAGCGGGAGTTGGTCTGGGCCAGAGAAGAAAGCCT 390

Qy  130 ThrProLysValAlaAsnLys 136
        |||     |||
Db  391 GAACCTTCAGTAGAAAAGAAG 411
```

```
RESULT 11
US-10-767-701-10044
; Sequence 10044, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 10044
; LENGTH: 1048
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
```

```
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS50751_1
US-10-767-701-10044

Alignment Scores:
Pred. No.:
Score:                        4.78e-05
Percent Similarity:           119.50
Best Local Similarity:        52.63%
Query Match:                  29.82%
DB:                           16.71%
                              17

US-09-701-618A-4 (1-141) x US-10-767-701-10044 (1-1048)

Qy   16 GluSerLeuThrAlaLeuGluSerSerLeuGlyAlaValAlaAspAspMetLeuLysThrMet 35
        |||:::||||:||:|||               :||     :::
Db  123 GAGACGCTGCCGCGGCTGCCGAATCC------------GTAGGGGGACCATCTTTCCATCTA 170

Qy   36 MetAlaValSerArg-----AsnGluLeuLeuGlnLysLeuLeuAspProLeuGlnIleAla 53
        :::||||               :::     ||:||     ||||||     :::|||
Db  171 CTCGCGCCGGCGGGGCGGCGGACCCCGACGCCGTGGCCGAGCTCCGCCCACTGCTCGGGCG 230

Qy   54 LysValAlaAspLeuValSerAlaThrTyrThrLeuAsnSerMetPheTrpValTyrLeuAlaThr 73
        :::            |||||     |||||    |||     ||||
Db  231 CGCGCGTTTCCTTGCCGTGGCGCAGGCCGCCGCCACTCTTCTCGGAGTTCGTTCGTTGTTTCGGGAGTGT 290

Qy   74 GlnGlyValAsnProLysGluHisProValLysGlnGluGluGlnLysGluArgIleArgValTyr 93
        |||||     :::            :||:||                :::
Db  291 TCGGGAATTGACCCTGACGAGCACCCCATCAGAAAGGAGAATCTGAAAGATTAAGCCTATCG 350

Qy   94 MetAsnArgValLys-----GluIleThrAspLys-------LysLysAlaAlaLys 108
        :::      |||               |||                ||||:::
Db  351 CAGGAGAAGTTAAAATCAATTTGAGGCCCTGGGACAAAGGACACAAGGCAACCACTTCGCCCTTCGCCCTACTACTACA 410

Qy  109 LeuAspArgGlyAlaAlaSerArgPheValLysLysAlaLeu 122
        :::      |||                      ||||
Db  411 ATAAATACACAAGCAGCACCAAGGTTTATTGGACACTCACTT 452
```

```
RESULT 12
US-10-437-963-24791
; Sequence 24791, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 24791
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT45530_29741C.1
US-10-437-963-24791

Alignment Scores:
Pred. No.:
Score:                        6.54e-05
Percent Similarity:           116.00
Best Local Similarity:        52.14%
Query Match:                  29.91%
DB:                           16.22%
                              17

US-09-701-618A-4 (1-141) x US-10-437-963-24791 (1-630)
```

RESULT 13
US-10-027-632-299522

; Sequence 299522, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 299522
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(476)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-299522

Alignment Scores:
Pred. No.:
Score: 103.00
Percent Similarity: 95.24%
Best Local Similarity: 14.41%
Query Match: 13
DB:

US-09-701-618A-4 (1-141) x US-10-027-632-299522 (1-476)

Qy 69 ValTyrLeuAlaThrGlnGlyValAsnProLysGluHisProValLysGlnGluLeuGlu 88

Db 352 ACCACTACAGTGAATACACAGGCAGCAGCAAGGTTCATTGGTCTCACTCATTG 402

Qy 107 ---AlaLysLeuAspArgGlyAlaAlaSerArgPheValLysLysLysAlaLeu 122

Db 292 AGCTTGTGGGAGGAAAAGTTAAATCGATTTGAGGACTGGGACAGGACCAGTTGCGCCCA 351

Qy 91 ArgValTyrMetAsnArgValLysGluIleThrAspLysLysLeuAla--------- 106

Db 232 TCGCGGTGTTCGGGAGTTGATCCGGATGAACACCCTATCAAGAAGAGAGTTTGAGAGGTTA 291

Qy 73 ThrGln----GlyValAsnProLysGluHisProValLysGlnGluArgIle 90

Db 172 GCGCGGCGGCGTTCTCGCCATGGCGAAGGCCGCGCCGCTCCCTCTTCGCAGGCAGGCAGCTC 231

Qy 53 AlaLysValAspLeuValSerAlaTyrThrLeuAsnSerMetPheTrpValTyrLeuAla 72

Db 112 ATGGTCGTGGCCGGGCGGGGAGGGACCCGAGCTTCCGCGCCGCGTCGCCGCGC 171

Qy 35 MetMetAlaValSerArgAsn------GluLeuLeuGlnLysLeuLeuAspProLeuGluGln 52

Db 52 GTGTCCGCGCCGAGGACACCCTTCGGCGGCCGGCCGAGTCCGTTCGGGGACCACCTTCGCCGAG 111

Qy 18 LeuThrAlaLeuGluSerSerLeuGlyAla---------ValAspAspMetLeuLysThr 34

RESULT 14
US-10-027-632-299522

; Sequence 299522, Application US/10027632
; Publication No. US20030240075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 299522
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(476)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-299522

Alignment Scores:
Pred. No.:
Score: 103.00
Percent Similarity: 95.24%
Best Local Similarity: 14.41%
Query Match: 15
DB:

US-09-701-618A-4 (1-141) x US-10-027-632-299522 (1-476)

Qy 89 Arg 89

Db 201 AGA 203

Db 141 GTTTATTTGGCAACCCAAGGAGAGTTAATCCTAAGGAACATCCAGTAAAACAGGAGTTGGTA 200

Qy 89 Arg 89

Db 201 AGA 203

RESULT 15
US-10-437-963-45565

; Sequence 45565, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.

Qy 141 GTTTATTTGGCAACCCAAGGAGAGTTAATCCTAAGGAACATCCAGTAAAACAGGAGTTGGTA 200

Qy 69 ValTyrLeuAlaThrGlnGlyValAsnProLysGluHisProValLysGlnGluLeuGlu 88

```
;  APPLICANT: Barbazuk, Brad
;  APPLICANT: Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE: 2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
;  SEQ ID NO 45565
;  LENGTH: 2823
;  TYPE: DNA
;  ORGANISM: Oryza sativa
;  FEATURE:
;  OTHER INFORMATION: Clone ID: PAT_MRT4530_48518C.1
US-10-437-963-45565

Alignment Scores:
Pred. No.:                          Length:          2823
Score:            0.0476            Matches:          34
Percent Similarity:  100.50         Conservative:     31
Best Local Similarity: 48.51%       Mismatches:       56
Query Match:       25.37%           Indels:           13
                   14.06%           Gaps:             5
                   17

US-09-701-618A-4 (1-141) x US-10-437-963-45565 (1-2823)

Qy      8 GluAspTyrProValGlnIleHisGluSerLeuThrAlaLeuGluSerSerLeuGlyAla 27
          ::|||||| ||| :::::: |  :::||| ::|||
Db   1708 GATGAGCGGCGCG---AAGCTGCTCGCGCCGCGTTGAAGGAGAGGGCCTCGACCTC 1764

Qy     28 ValAspAspMetLeuLysThrMetMetAlaValSerArgAsnGluLeuLeuGlnLysLeu 47
          |||          :::                    |||
Db   1765 GTCACCGGCCAAGGTCAAGGCCGCTCACCCGCAAGGTCAAGAAGAACCAGCTGCCGACGGCG 1824

Qy     48 AspProLeuGlu------GlnAlaLysValAspLeuValSerAlaTyrThrLeuAsnSer 65
          ||| :::         ||||||| ||:: :::|||
Db   1825 GATGGGGATTGGGTACCTGGAGGCAAAAGCACCACCTGCTGCTTGAGCTACTGGCCAGGACATC 1884

Qy     66 MetPheTrpValTyrLeuAlaThrGlnGlyValAsnProLysGluHisProValLysGln 85
          :::||||||||:::::: ||||||| ::::::          |||          :::
Db   1885 GTCTACTATCTCTCCCCAAGGCGCGAACGGGGCTTCTCGGTGGAAGAGGTCACCCCGTGCGG 1944

Qy     86 GluLeuGluArgIleArgValTyrMetAsnArgValLysGluIleThrAspLysLysLys 105
          |||      |||||||:::::::: :::||| :::|||          |||
Db   1945 AGCCTTGTCGAGATCAGGCGTGTTCTTGGAGAAGATTCGTCCCATA---GACAAGAGATG 2001

Qy    106 AlaAlaLysLeu-----------------------AspArgGlyAlaAlaSerArgPheVal 118
          :::||||                             |||||||||||          |||
Db   2002 GAGTACCAAATCCAGAAGCTCACCACGCCGCGGATAGCGGAGCTGCCCAGGAGGAAGGTG 2061

Qy    119 LysLysAlaLeuTrpGluProLysGlyLysSerThrProLys 132
          |||     ||| |||||| ||||||
Db   2062 CTAAATGCA------GAAGCAAAGAGTAAGGATCAGCCTAAG 2097
```